Psychological Bulletin

HARRY HELSON, Editor
Kansas State University

CONTENTS

Published Bimonthly by the
American Psychological Association

VOL. 59, NO. 1 JANUARY 1962

The *Psychological Bulletin* contains evaluative reviews of research literature and reviews of research methodology and instrumentation in psychology. This Journal *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to the Editor, Harry Helson, Department of Psychology, Kansas State University, Manhattan, Kansas.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50–60), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

Reprints. Fifty free reprints are given to contributors of articles and notes.

Helen Orr
Acting Production Manager

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D.C. Address changes must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin

and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Washington, D.C., and at additional mailing offices. Printed in U.S.A

VOL. 59, NO. 1 JANUARY 1962

Psychological Bulletin

RESEARCH WITH THE WECHSLER INTELLIGENCE SCALES FOR ADULTS: 1955-60[1]

WILSON H. GUERTIN
University of Florida

ALBERT I. RABIN
Michigan State University

GEORGE H. FRANK AND CLAYTON E. LADD
University of Miami *Indiana University*

Two important events took place since the bulk of the material for our previous review (Guertin, Frank, & Rabin, 1956) was gathered and organized. The first is the publication of the manual for the revised WB,[2] known as the WAIS (Wechsler, 1955); the second was the appearance of a new, rewritten, and reorganized edition of Wechsler's *Adult Intelligence* (1958).

Although the manual was mentioned in our previous review, the new test it introduced—the WAIS—had not yet become the popular instrument it is today. It seems to be replacing the old WB as a research tool and as a clinical and assessment device for many good reasons. For reviews of the WAIS see Buros (1959). The present review covers work done with both instruments for it spans a period of transition.

In the closing summary of the previous review, we expressed the hope for "the creation of a newly standardized instrument, similar in structure to the WB, but not suffering from the numerous weaknesses." The WAIS, in many respects, is the answer to this wish. A fairly rich harvest of research with this method is critically considered in the following pages. It may be added, in agreement with Wittenborn (1957), that:

> There is a refreshing trend away from gross empirical validations which required that tests predict the diagnostic decisions of psychiatrists or psychologists. Instead, there seems to be an emphasis on the conceptual validity of the procedures employed in assessment (p. 331).

The general outline of the present review and its organization are quite similar to our previous reviews (Guertin et al., 1956; Rabin, 1945; Rabin & Guertin, 1951). The amount of material covered under each rubric differs however, for some

[1] Through July 1960.

[2] The abbreviation, WB, will be used throughout to indicate the Wechsler-Bellevue Intelligence Scale, Form I. Form II will be designated WB II, while WAIS signifies the Wechsler Adult Intelligence Scale. The names of the subtests also appear in abbreviated form throughout the paper. The single letters I, C, A, D, S, and V stand for the Verbal subtests: Information, Comprehension, Arithmetic, Digits, Similarities, and Vocabulary, respectively. The two-letter combinations PA, PC, OA, BD, and DS correspond to the following Performance subtests: Picture Arrangement, Picture Completion, Object Assembly, Block Designs, and Digit Symbol, respectively. FS, VIQ, and PIQ stand for Full Scale, Verbal IQ, and Performance IQ, respectively.

currents have run dry, while previous trickles have expanded markedly. The bibliographical coverage is selective in view of differences in relevance, quality, and significance of the various researches reported in the literature.[3]

As a Measure of Intelligence

Reliability

An inspection of Wechsler's tables (1958, pp. 102–3) suggests that the WAIS IQs and verbal subtests are slightly more reliable than comparable WB IQs and subtests, but that the performance subtests (possibly excepting DS) have about the same reliability coefficients on both tests. Perhaps this indication of increased reliability with the WAIS has curtailed the number of studies reporting test-retest or split-half reliabilities for this test as only one has been published thus far. Over long periods ranging from 1 to 5 years and using bright "normals" (Bayley, 1957) or psychiatric patients (Armitage & Pearl, 1958), the WB has yielded test-retest correlations similar to those found in earlier reliability studies, i.e., .77–.95.

Coons and Peacock (1959), using 24 mental hospital patients, obtained test-retest correlations for all three WAIS IQ scores of .96 or better, and the standard errors of measurement were consistent with those obtained with the standardization sample. From this it was inferred that:

> IQ changes on retest with different examiners of more than 6 points can be attributed with reasonable confidence to changes in the mental state of the patient.

Yet, the practice effects or at least increments in IQ scores at the time of the second testing were 2.6, 8.6, and 5.0 points for VIQ, PIQ, and FSIQ, respectively. Consequently, the quoted inference needs a qualification, such as "after appropriately adjusting for practice effects." Test-retest differences were not only greater but also more variable for the PIQ than for the VIQ or FSIQs; thus, it was concluded that "the Verbal scale is a better indicator of the level of the original Full Scale performance than is the Performance Scale IQ." At the subtest level, the test-retest reliabilities are generally higher than the split-half reliabilities reported in the WAIS manual (1955). D had the lowest reliability of all subtests with a .84; the other Verbal subtests (excepting C with a .89) were .94 or better. The Performance subtests averaged .88, suggesting to the authors that the Verbal subtests are more reliable than the Performance subtests; however, one should remember that the practice effects were much more variable on the Performance subtests, which would reduce the test-retest reliability coefficients.

Comparative Validity

WB II and WISC. Earlier comparisons of the WB and WB II disclosed that practice effect was appreciably greater when the WB II was administered first. Thus, a very interesting and mystifying phenomenon confronted and worried Wechsler workers until Barry, Fulkerson, Kubla, and Seaquist (1956) failed to find a significant interaction between

[3] A supplementary bibliography along with the references covered by this review aims at complete coverage of research articles employing the adult Wechsler scales. This supplementary bibliography has been deposited with the American Documentation Institute. Order Document No. 6843 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

practice effect and the form of the WB administered first. Furthermore, they reported lack of equivalence between forms for entirely different subtests (S, OA, and DS) than reported by earlier workers. Their equivalent form reliability coefficient of .71 is consistent with earlier findings and is rather high since their range of talent (intelligence) was only half that of an unrestricted sample.

Findings of earlier comparisons between the WISC and WB were confirmed by Price and Thorne (1955). Their sophisticated statistical analysis of data disclosed a slightly lower WB FSIQ and VIQ, while PIQ was slightly higher than for corresponding WISC scales. Correlation between the FSIQs was very high for their 11.5-year-old sample (.89) and moderate for their 14.5-year-old sample (.78), but range of talent was considerably lower in the older group.

WAIS. Cole and Webela (1956) reported a comparison of the WB and WAIS, but their restricted range of talent and incomplete counterbalancing of the form of the test with order of administration prevent any findings from being more than suggestive. Goolishian and Ramsay (1956) also were interested in the equivalence of the new WAIS and the WB, so they studied the two arrays of test scores in their hospital files. While the design employs different subjects for the two test scores, thus permitting the operation of sampling biases, the investigators employed a large *N*. They failed to find the extreme differences noted by Cole and Webela, but five subtests showed significant differences between the two tests. Neuringer's careful study (1956) showed FSIQ and PIQ were higher for the WB, a finding echoed in a more subjective report by Sinnett and Mayman (1960). Dana's results (1957a), based upon a study of only the Verbal scales, revealed no significant differences for any of the subtest comparisons, a finding that is quite different from that of Cole and Webela. Then, in support of the large differences between forms found by Cole and Webela; Karson, Pool, and Freud (1957) reported significant differences for five subtests, also providing confirmation of some of the Goolishian-Ramsay findings. Light and Chambers (1958) found, with defectives, that the WAIS, VIQ, and FSIQ were significantly higher than for the WB. Correlation of the FSIQ was .77 for their restricted range of talent sample. Garfield (1960) found BD to be ninth in WAIS subtest order of difficulty as compared with third place for the WB BD.

It would appear that the only consistent finding with samples of average or higher intelligence is higher scores on BD, DS, PIQ, and FSIQ for the WB; and there is little agreement as to which of the verbal subtests are lower for the WB, if any. Only Neuringer's study (1956) had all the necessary features of appropriate range of talent, sufficient *N*, unbiased samples, and appropriate counterbalancing to test the equivalence of the WB and WAIS. After correcting for range of talent, Neuringer's correlations for VIQ, PIQ, and FSIQ, respectively, were .89, .44, and .77—hardly satisfactory for "equivalent form" reliability.

Other tests. Sines (1958) reported correlations of .77, .78, and .79 between the Shipley-Hartford and the WB FS scores for three samples and provides regression equations for predicting WB FSIQ from the Shipley. Three tests from the Army Classification Battery correlated .60 to .81 with the WB FS scores (Montague,

Williams, Lubin, & Geiseking 1957), while Murphy and Langston (1956) obtained a .83 between the WB FS score and the Army Classification Battery, Area Aptitude I Test. Higher correlations between the Revised Beta and WAIS (.81 and .83 for Negro and white prisoners) were found by Panton (1960).

Sterne (1960) reported a correlation of .84 between the Ammons Full Range Picture Vocabulary Test (FRPV) and the WAIS FSIQ for a sample of older medical patients. Allen, Thornton, and Stenger (1956), using college students with a markedly restricted range of talent, obtained a correlation of only .46 between the FRPV and the WB FSIQ. Fisher, Shotwell, and York (1960) found correlations between FRPV and various WAIS scores ranging from .36 to .79 with defectives. Borgatta and Corsini (1960) reported correlations between WAIS FS scores and four forms of their Quick Work Test of .75 to .83, with the observation that coefficients are attenuated by reduced range of talent. Rabinowitz (1956) compared the Kent EGY with the WB FSIQ and found a correlation of .69 for hospitalized psychiatric patients with a normal range of intelligence.

Those interested in Raven's Progressive Matrices often use the Wechsler for comparative purposes. Hall (1957a) found a .72 correlation with the WAIS FS scores, while Stacey and Gill (1955), working with the restricted range of talent found in samples of adult defectives, reported a correlation of .68 with the WB FSIQ. Urmer, Morris, and Wendland (1960), and Moya-Diaz and Matte-Blanco (1953–55) also studied the matrices and Wechsler scores. The latter found the tests fairly equivalent but noted that anxiety and cultural factors were more important determinants of WB scores than for scores on the matrices. Confirming this, Levinson (1959) employed a sample of 80% foreign born with two age ranges. Matrices scores correlated with the WAIS FSIQ .65 for his 60–69 year olds and .40 for his 70–79 year olds. As expected, he found a negative correlation between WAIS performance and age, which was greater in the older group. Had he used WAIS weighted scores instead of IQ, he would have obtained higher and more appropriate correlations with the matrices.

Hall (1957b) found the WB FS scores and Wechsler Memory Scale correlated .75 and concluded there was a large overlap in what the two tests measure. Strong (1959) found a mixture of WAIS and WB FSIQs correlated .63 with the Ohio Literacy Test for psychiatric patients. One would expect a higher correlation for weighted score than IQ since the Ohio Literacy Test has no correction for deterioration with age.

Summary. The studies reviewed in this section, when compared as a whole with those covered in the last review, are very disappointing. Not only have the investigators failed to learn from others' mistakes, but there seems to be little tendency to design critical and conclusive studies to resolve conflicting findings reported earlier.

Range of intelligence in the sample is often ignored, frequently not reported, and only one correlational study employed a correction for restricted range of talent (intelligence).

It seems useless to remind investigators that equivalence between tests depends upon *both* correlation and differences in mean scores, but we would be remiss were we not to repeat this again. Somewhat encouraging is the tendency seen to use the more sophisticated approaches of analysis

of variance and regression equations for specifying IQ.

Short Forms

The new WAIS has given a fresh impetus to studies involving short forms. In an early article concerning the WAIS, Doppelt (1956) decided upon the tetrad short form (A, V, BD, and PA) consisting of the two subtests which correlated most highly with their respective scale scores in Wechsler's standardization population. Doppelt presented a regression equation method of computing the FS score which was compared by Himelstein (1957b) with simple prorating. Himelstein found the total scores computed by the two methods correlated .99 and since the means were identical, concluded that the clinician may feel free to use either method.

The Doppelt article was the partial stimulus for a rash of studies (Clayton & Payne, 1959; Fisher & Shotwell, 1959; Himelstein, 1957b, 1957c; Olin & Reznikoff, 1957; Sines & Simmons, 1959; Sterne, 1957; Whitmyre & Pishkin, 1958) reporting the application of Doppelt's WAIS short form to patient populations and generally concluding that this abbreviated scale provided about as valid an estimate of the FS score for heterogeneous psychiatric subjects as for the standardization subjects. While correlations range from .92 to .97, it must be remembered that they are exaggerated since they represent correlation of parts with the whole. Findings for samples with restricted range of talent gave lower short form-FS correlations for homeless men (Levinson, 1957), mental defectives (Clayton & Payne, 1959), and students (Allen et al., 1956). Both Levinson's and Himelstein's comments (1957a) ignore the constricting effect of the reduced range of talent in Levinson's sample on the size of the obtained correlation, which, when corrected, rises from .87 to .92. Sterne (1957) similarly found a lower correlation with organics but the obtained coefficient is highly unreliable with $N = 12$.

Using a similar formula to that developed by McNemar for the WB, Maxwell (1957) determined the correlation of all possible two, three, four, and five subtest combinations with the WAIS FS for the 300 subjects in the 25–34 age group of the standardization population. She concluded: (*a*) that the accuracy of abbreviated scales is a function of the number of subtests included; (*b*) that while short verbal scales are generally better than performance scales as predictors of FS scores, a combination of both verbal and performance subtests is best; (*c*) that the best WB and WAIS abbreviated scales are not composed of the same subtests; and (*d*) that WAIS short forms are more highly correlated with the FS than are the WB short forms. The last conclusion was challenged by Howard (1958) who contends McNemar made an error and underestimated the correlations between WB abbreviated scales and FS. Howard (1959) also reported finding higher WAIS short form-FS correlations in a group of heterogeneous psychiatric patients than Maxwell found in the standardization sample, but he recognized that "the differences appeared to result from the greater variance of the patient sample."

Three studies within the last 5 years have considered the usefulness of WB II abbreviated scales for employee selection (Sloan & Newman, 1955), with alcoholic outpatients (Schneyer, 1957), psychotics and students (Caldwell & Davis, 1956).

Special Populations and Applications

Intelligence as a function of age. Bayley (1957) concerned herself with the growth of intelligence between 16 and 21 years of age in an extension of the now famous Berkeley Growth Study. In general, subjects improved with each testing regardless of intellectual or educational level. Certain individuals, however, appeared to have reached their asymptote by 16 or 18 while others continued to develop until 21 or older. Although acknowledging possible practice effects, Bayley did not feel this totally accounted for the increments in performance.

Concerned with the encroachments of old age in a randomly selected probability sample in Delaware, Whiteman and Jastak (1957) administered three subtests of the WB to 1,980 persons and found little decline with age on C, moderate decline on PC, and marked decline on DS beginning at age 35. These differential deficits in performance accruing with age were interpreted as "a decline in certain group and specific factors—conative, perceptual, and motoric in nature—rather than as a decline in general intellectual ability *per se.*" Similar interpretations of the WAIS standardization data were made by Doppelt and Wallace (1955) and Wechsler (1958). Comparing the WB standardization population with the WAIS standardization population, Wechsler (1958) noted that the best overall WAIS test scores occurred in the 25–29 age interval rather than the 20–24 age interval found for the WB standardization. Also, the general rate of decline was said to be less for the WAIS than for the WB up to age 50.

Doppelt and Wallace (1955) found that allowing the elderly subjects unlimited time made very little difference in their scores. The WAIS standardization population scores began to decline with aging much sooner, and decrement was much more marked on the Performance subtests than on the Verbal ones. The WAIS Verbal subtests hold up fairly well until about 70 years of age at which time all subtest performances decline rapidly with age. Eisdorfer, Busse, and Cohen (1959) questioned the representativeness of the WAIS Kansas City aged sample (Doppelt & Wallace, 1955); however, when 162 volunteer subjects from the Piedmont section of North Carolina consistently (82%) manifested a superiority of VIQ over PIQ. This Verbal superiority remained even when sex, race, socioeconomic, intelligence, and mental health differences were analyzed separately. It is noteworthy that the VIQ-PIQ discrepancy for the entire sample is more attributable to an elevation of the VIQ (106.5) above the norm than to a depression of the PIQ (98.5). It may be that their *volunteers* show a greater relative elevation of verbal skills than the WAIS standardization sample. Loranger and Misiak (1960) found DS performance of a group of aged females comparable to that of the Kansas City standardization sample.

Sex differences. In the WAIS standardization population there were consistent but negligible differences in Verbal Performance and FS scores in favor of the males (Doppelt & Wallace, 1955; Wechsler, 1958). Eight of the 11 subtests showed significant sex differences with men doing better on five (I, C, A, PC, and BD) and women better on three (S, V, and DS). Apparently the rise and fall of the Mental Deterioration Index has had little effect on Wechsler's habit hierarchy, for he now proposes a new "WAIS masculinity-femininity (MF) score" composed of the F total (V+S+DS) subtracted

from the M total (I+A+PC). In the Plant and Lynd (1959) norms for 361 college freshmen there were no statistically significant sex differences on any of the WAIS IQs but subtest scores were not reported. In the Berkeley Growth Study (Bayley, 1957), males were superior on the Verbal scale, while females were higher on the Performance scale; however, there was no evidence for an earlier intellectual maturation of females. An unpublished thesis by Miele (1958) deals with sex differences on the WAIS.

Educational and vocational applications. The general intellectual level of college students has long been of interest. Plant and Richardson (1958) recently reported a mean WB FSIQ of 116.5 for college freshmen volunteers. Wechsler (1958) reported a very similar mean. Plant and Lynd (1959) found correlations of Verbal, Performance, and FS WAIS weighted scores with grade point average for the freshman year were .58, .31, and .53, respectively, which were as good or better than similar correlations for the ACE. Their normative data reveal an expected restriction in range of talent. The WB VIQ for engineering students has been reported (Wechsler, 1958) to be not only superior to the PIQ but also more highly correlated (.41 vs. .08) with college grades. Weisgerber (1955) concluded that Diamond's factor analytically based scoring method designed for vocational counseling with the WB was not as useful as the VIQ for predicting academic success of engineering students. At an even higher educational level, Holt and Luborsky (1959) have indicated their surprise at finding the WB VIQ to be one of the better predictors of performance in psychiatric residency training in spite of the test's ceiling. Correlations between the WB VIQ and supervisor-peer ratings on diagnosis, therapy, administration, management, and overall competence ranged from .27 to .47; even the correlations with empathy, interest, sensitivity, firmness, etc. were in the .30s.

A very interesting and thorough study of the relationship between intelligence (WB) and rated creatively in 64 chemists engaged in industrial research has been reported by Meer and Stein (1955). Not too surprisingly, when the entire group was considered there were generally positive findings although not always significant relationships among education, intelligence, and creativity. Their probing analysis, however, led to the tentative conclusion that:

> Where equal opportunity is available higher IQ scores beyond a certain point [approximately Percentile 95] have relatively little significance for creative work.

Considering the role of intelligence in managerial positions, Balinsky and Shaw (1956) found their unique sample had a higher WAIS VIQ (125) than PIQ (117) and, after correlating the IQs and subtest scores with overall performance ratings by superiors and peers, concluded that:

> Apparently verbal intelligence and especially arithmetical ability are important factors in the performance of the executive personnel.

While one might argue with the authors' phraseology—"important factors"—since the data indicated only one (A) of the 11 subtests yielded a significant correlation, the VIQ-performance rating correlation of .32 was significant at the .05 level. Another study, by Dunnette and Kirchner (1958), provides some confirmation of this relationship between intelligence and managerial effectiveness.

Cultural influences, translations, and ethnic groups. Bloom (1959)

recently compared 67 student nurses in Missouri with 67 in Hawaii using the V and PC subtests of the WAIS. The Missouri nurses obtained higher scores on both subtests (significant at .01 level only for V), and seven of the eight hypotheses about ecologic difficulty of PC items were confirmed. In a similar fashion, Breiger (1956) compared the WB PA performance of 30 United States Caucasians, 20 Nisei, and 10 German refugees. The three groups matched on IQ, education, urban-rural residence, and bilingualism, scored approximately the same on this subtest when evaluated in the usual manner, but a content analysis of stories related to their own arrangement of the "Flirt" and "Taxi" items revealed marked differences. Significantly more Caucasians than Nisei project romantic implications into the Flirt sequence and abnormal sex behavior into the Taxi arrangement. Sullivan (1957), in testing 15 and 16 year olds in Newfoundland, found rural subjects were handicapped on the WB.

Numerous applications of the WB and WAIS to foreign populations are evident during this 5-year period, and most of these investigators have found it necessary to make modifications of varying degrees to the test to correct for cultural biases. New translations of Wechsler's third edition have been made into French (Chagnon, 1955) and German (Wechsler, 1956). The WB has been translated into Danish and tried out with institutional cases (Mogensen, 1958 unpublished). Italian prisoners have been tested (Lazzari, Ferrecuti, & Rizzo, 1958). Priester (1957), and Priester and Kukulka (1958) presented a method of comparing HAWIE (German WAIS) subtest scatter with Wechsler's diagnostic signs. He also compared the HAWIE with the HAWIK (German WISC) and the Binet-Bobertag, finding them sufficiently comparable to be considered parallel tests. Cultural aspects of the WAIS in Canadian subjects (Hopkins, 1957) and in British mental patients (Robertson & Batcheldor, 1956) have been reported. The latter authors concluded the British subjects were better on literary and poorer on scientific I and V items than the American standardization sample; accuracy rather than speed characterized the British approach.

More directly to the point were a series of discerning articles by Levinson (1958, 1959) who expounds the thesis that reliable and valid differences between VIQs and PIQs are not necessarily the result of pathology but may reflect the deviant values associated with specific subcultures. He substantiates his case by citing the WAIS scores of 64 Yeshiva University students who had been indoctrinated with the traditional Jewish cultural values that place great stress upon verbal accomplishments and discount manual skills. This group obtained a mean VIQ of 125.6 but a mean PIQ of only 105.3, with 97% of the subjects having a higher VIQ than PIQ.

A well-designed investigation comparing the youngest WAIS standardization group with 100 Navaho Indians of comparable age, sex, education, occupation, and rural-urban residence (Howell, Evans, & Downing, 1958) afforded a striking contrast with the studies of Jewish students. The Navaho group obtained a VIQ of 84.0 and a PIQ of 95.4, which were significantly lower than those of the standardization group. Another group, however, which also stresses manipulative skills more than verbal accomplishments, the Southern Negro, showed a slight and nonsignificant tendency for the WB II VIQ to be higher than PIQ

(Davis, 1957). This was true for both his mental patients with various diagnoses and hospital employees, but perhaps most significant were the absolute levels (mean FSIQ 68 for the employees and 67 for the patients). A question concerning educational background of these groups arises, and a supplementary investigation indicated that both groups compared favorably on amount of education with the 1950 census figures for nonwhites in Florida. Scarborough (1956) compared 40 venereal diseased patients with 118 control subjects in a complex, poorly designed study and derived inconclusive results. His findings suggest that Southern Negroes do less well on the WB (IQs $\approx$ 80) than Southern whites (IQs $\approx$ 90) and that the patients of either race do almost as well as their own control group. The Negro subjects in this and in the Davis study did relatively well on OA but poorly on D and DS. Just why Scarborough's Negro subjects from Georgia should average almost 13 IQ points higher than Davis's Negro subjects from Florida is puzzling.

Some very interesting information about the intellectual distribution of 3,594 unwed mothers placing their children for adoption in Minnesota was provided by Pearson and Amacher (1956). The mean IQ was 100.19 with a standard deviation of 18.36. Although approaching a normal distribution, there were fewer cases than expected between IQ 83 and 91. The authors hypothesized that these deviations were due to a greater proportion of mothers falling at the extremes of the intellectual continuum placing their babies for adoption because of necessity or social pressure, while dull normal mothers more commonly keep and rear their illegitimate children. It is noteworthy that "repeaters" obtained a mean IQ of 93.3.

Summary. Intellectual growth, as defined by improved Itest performance on the WAIS continues in our culture until 25–30 years of age, but wide individual differences exist in the age of maturation ranging from the early or middle teens to the late twenties or older. Shortly after the intellectual peak, however, aging makes its first encroachments upon perceptual and psychomotor tasks; only considerably later does it appreciably affect verbal skills. Whether Wechsler's (1958, p. 143) conceptual distinction between "intelligence" and "wisdom" (defined by reference to the ability of the old sage to cope with life's problems) is useful remains to be seen, but an obvious implication is that a test for each concept is needed at least to evaluate the hypothesis that both are worthwhile. Although sex differences have been demonstrated fairly consistently on certain subtests, IQs are usually comparable. In addition to age and sex, a variety of environmental influences, such as subcultural background and values, education and vocational history, socioeconomic conditions, etc., may produce diverse and dramatic effects upon intelligence test scores.

Thus, the conclusion of this section remains essentially the same as in previous reviews although valuable new data has been added, namely, that a number of variables besides pathology affect Wechsler performance and consequently must be controlled or accounted for in adequate analyses. Clearly, no one can criticize Dunnette and Kirchner's (1958) plea for validity studies in specific vocational stiuations instead of reliance upon the assumed intrinsic validity of a test.

Refinements and Critiques

Administration and scoring. In contrast to the last review, only one

paper is concerned with item order and difficulty of the WB. Rubin-Rabson (1956) points out the time boundedness of previously established item orders and observes the undesirable "tendency for items to cluster in groups of similar difficulty, [and] an abrupt augmentation of difficulty from group to group."

Two important investigations of the effect of administration and test taking attitudes were reported. Masling (1959) slyly coached his accomplices, as he appropriately calls them, for "warm" and "cold" roles to be played when tested by unsuspecting experimenters. Utilization of some memorized answers, taped sessions, and a set of judges demonstrated that the warm role enhanced the score in three ways: experimenters used more reinforcing comments, gave more opportunity to clarify and correct answers, and scoring was more lenient toward the warm subjects. However, these statistically significant differences were small.

Nichols (1959) manipulated ego involvement and success experience for college students taking the WB. He concludes:

> differences in test taking attitude on the part of the *S* and minor differences in testing procedure on the part of the *E* do not materially affect intelligence test scores. [He adds this important caution] However, since the subjects used in this study were all intelligent students who are used to taking tests and doing their best, the results may not be directly applicable to clinic and hospital groups.

We would add: or to children.

The effects of a trusting or skeptical attitude in student nurses upon the WAIS S and PC subtests were investigated by Wiener (1957) who hypothesized that a distrustful attitude would increase the "no similarity" or "nothing missing" responses and thus interfere with performance on these subtests. The attitudes were measured by a questionnaire, and distrustfulness was also presumably reinforced or induced by special instructions. The more distrustful students on the questionnaire displayed a stronger tendency to make the predicted distrustful comments and were lower on both S and PC subtests. The experimental instructions, however, did not depress the subtest score but did increase the number of comments suggestive of distrust. The results are interesting and suggestive, but it should be noted that the *N* was small and that only difference scores (S−V and PC−V) were reported.

Guertin (1959) found that various, controlled background noises had no effect on D performance with a group of chronic psychotics. But, again, distraction would be more likely for subjects who maintain more interest in their surroundings, so generalization about the unimportance of noise during D administration is most hazardous. Blackburn and Benton (1957) suggest a more reliable administration and scoring procedure for D. They present reliability data from several populations and give conversion tables. Briggs' study (1960) is reassuring in that only DS results were appreciably affected when the subject was forced to manipulate with his nondominant hand. Plumb and Charles (1955) studied scoring disagreements to C responses and found that experts as well as graduate students disagreed significantly. Olin (1958) presents tables taking into account the subject's age group when prorating IQ. Clinicians making prorations of IQ in the aged from short forms should note that unless Olin's procedure is followed, they are introducing appreciable error in estimating IQ.

Factor analyses. Davis (1956) derived 10 factors from the WB subtests, many more than previously reported. His use of a narrow range

of talent emphasizes test or methods factors as opposed to trait factors and increases dimensionality. Saunders (1959a) observed that Davis used a nonuniform procedure in obtaining intercorrelations that also could account for the unexpectedly large number of factors. Not stopping with criticism, Saunders devised a crucial test of the dimensionality of the WAIS. He divided the subtests into odd-even to increase the number of variables, thereby avoiding restriction on the number of factors forthcoming. From this model study he concludes:

> The results are consistent with the efforts of some clinical psychologists to interpret the Wechsler "psychogram" as a personality measure provided attention is given to individual items of C and PC. Results are also consistent with prior factor studies of the Wechsler which have found only three to five [group] factors.

Cohen (1957b) found factors on the WAIS similar to those obtained earlier on the WB. Besides a strong general intellectual factor he found a verbal comprehension, a perceptual organization, and a memory factor. Findings based upon four age groups lead him to conclude:

> This evidence is contrary to Garrett's "differentiation hypothesis," which suggests a sharp reduction in the importance of the general factor by the late teens.

He notes the exception that the memory factor tends to supplant much of the general factor in the old age group. He feels that the rather low amount of subtest specificity encountered helps account for disappointing outcomes with pattern analysis. Zwart and Houwink (1958) also found three WAIS subtest factors, two of which corresponded closely to Cohen's factors.

Saunders (1960b) reanalyzed his own WAIS data to study the factors involved in PC subtest items. Findings are interesting and important to the WAIS user since three distinct, clinically meaningful factors emerged. In another reanalysis, Saunders (1960a) found *six* factors were necessary to account for I and A responses. The complexity of I and the inappropriateness of an over-all subtest score for pattern analysis is illustrated by the appearance of five factors involved in this single subtest. Three factors underlie the A subtest.

Subtest rationale. Saunders (1959b) discusses the rationale of the Wechsler subtests in terms of clinically derived hypotheses that are consistent with early statistical findings. Cohen (1957a) similarly discusses WAIS subtest rationale in the light of his factor analytic findings.

Levine (1958) concentrated on S and separated out the "not alike" responders. He found they had a lower mean IQ and he discusses the theoretical implications. In another study Levine, Glass, and Meltzoff (1957) separated out the "*N*" reversers on DS and found they too were less intelligent and "cognitive inhibition time" (capacity to delay a response) was poorer than for controls.

Matarazzo and Phillips (1955) were interested in the relationship between manifest anxiety score and DS performance. They believed a nonmonotonic function best explained their data. When Goodstein and Farber (1957) examined the relationship between manifest anxiety and DS score, they included a very anxious group to extend the range of anxiety upward in the hope of clarifying the nature of the relationship, but no significant relationship of any kind could be recognized.

Heilbrun (1960) calculated the intercorrelations of four immediate memory tests including WB D for brain damaged and control patients. All intercorrelations for both groups were significant (ranging from .26 to

.62), suggesting a general memory factor but, nevertheless, of such restricted magnitudes as to dictate "considerable caution" in deriving conclusions regarding an individual's general memory functioning from only one test.

Summary. This section represents interest constructively directed at how the Wechsler works and what can be done to improve it; thus, it is disappointing to see that there are somewhat fewer articles covered than in the previous review. However, the quality of the articles is generally good. Cohen (1957b) continued to contribute methodologically by using age groups in factor analytic design. Saunders (1959a) has provided us with a first look at the specific and group factor structure of the Wechsler. His factor analyses of subtest items has been most productive and we look forward to further reports of these findings and the time when he will bring forth an up-to-date rationale for all the subtests. Nichols' (1959) manipulation of ego involvement and success experience provides important information and needs to be extended to other populations.

The Wechsler as a Diagnostic Aid

Personality Variables

Anxiety. In most studies the criterion measure for anxiety was the Taylor scale. Using a wide variety of subjects, such as psychiatric aides compared with outpatient state hospital patients; high and low anxiety groups of college undergraduates, or medical compared with psychiatric VA patients, Dana (1957b); Goodstein and Farber (1957); Mayzner, Sersen, and Tresselt (1955); and Matarazzo (1955) found no consistent relationship between the Taylor scale and Wechsler scores. Siegman (1956) found that Taylor scale anxiety was associated with lowered performance on timed subtests only. However, using a college population Calvin, Koons, Bingham, and Fink (1955) found a consistent relationship between scores on the Taylor scale and diminished efficiency on such WB items as FSIQ, VIQ, V, I, D, A, BD, and OA. Not using the Taylor criterion, Griffiths (1958) assumed induction of anxiety in a group of college freshmen exposed to an experience of failure in a testing situation. As compared to controls, significantly lower performance was observed on D and I but not on A, OA, or DS.

Kerrick (1955) found that anxiety disrupted over-all performance of Air Force trainees on the WB, whereas in a similar study, Mayzner, Sersen, and Tresselt (1955) failed to observe such impairment with college students. Mayzner et al. hypothesized that the differences in the findings between the two studies might be attributable to the appreciable anxiety of Kerrick's Air Force trainees, who realized the greater relevance of the test results to their future careers in service, as compared with the college subjects.

Miscellaneous. Tallent (1958) was unable to support the clinical interpretation that ninth grade boys saying "yell fire" to the C "theatre" item are impulsive behaviorally as judged by their teachers. Of course, the negative results might equally well indicate that teachers have little recognition of their students' impulsiveness. Of related interest is the finding that "ego delay function," as measured by Barron M-threshold inkblots, time estimation, and Stroop Color-Word Test, was correlated with WBIQ and D (Spivack, Levine, & Sprigle, 1959).

The WB has also been evaluated as a predictor of continuation in psychoanalytically oriented therapy (Hiler, 1958). Patients remaining in treatment for at least 20 sessions averaged about 10 points higher in

IQ (mean IQ 112) and did better on S but poorer on D and DS relative to the other subtests than the patients discontinuing within five sessions. McReynolds and Weide (1960) reported dramatic changes on DS following prefrontal lobotomies, but the subtest given preoperatively was not predictive of the degree of psychiatric improvement postoperatively.

Investigations of Diagnostic Value

Several studies regarding the general diagnostic usefulness of the WB have appeared to reinforce our cautious, skeptical approach to the clinical application of tentative relationships between test results and psychiatric condition. Frank (1956) correlated and factor analyzed the subtest scores of 60 subjects from nine diagnostic groups which, in a previous analysis, appeared homogeneous in subtest scores. Only two unrotated factors were isolated: VIQ and PIQ. The conclusion was that "the WB does not yield significant data as regards psychiatric diagnosis, and continues to sort subjects in terms of intellectual factors only." Cohen (1955) submitted WB profiles of 300 male veteran patients diagnosed as psychoneurotic, schizophrenic, or brain damaged to seven experienced clinical psychologists and had them attempt to classify each case. Only one of the seven psychologists correctly classified a significant number (132) of the 300 patients and only two others had above-chance success in the diagnosis of a single diagnostic group which in both cases was the brain damaged group. The judged classification correlated with the neuropsychiatric diagnosis is between .13 and .22, which was deemed far too small to be of use clinically. It was concluded that there is some nonchance relationship between the WB pattern and the clinical diagnosis but that this relationship is detected by only a few clinicians and even then to only a degree having little practical value. Despite these and earlier studies, some clinicians continue to use the test diagnostically with little hesitation.

Almost at the other extreme, however, are the clinicians who discount or disregard the possible influence of emotional or environmental factors upon IQ scores. For example, Garfield and Affleck (1960) reviewed 24 cases committed to an institution for the retarded but later judged not mentally defective and found the IQ played an important role in the commitment proceedings. In most of these cases serious emotional problems, deprived environments, or uncooperativeness existed but were neglected by the psychometrist who proceeded to write with finality a report diagnosing mental deficiency and indicating a poor prognosis. The gross misinterpretations and misuses of the IQ described in this article should arouse some concern over maintaining acceptable standards for practicing psychometrists.

Rabin, King, and Ehrmann (1955) found long-term schizophrenics were lower than normals and short-term schizophrenics on the WB Vocabulary. Normals and short-term schizophrenics did not differ significantly. Characteristics of the stimulus word also affected the level of communication; thus, it seemed that the possible effects of chronicity, severity of the pathology, type of verbal material, and scoring system should all be considered in investigations involving verbal behavior of schizophrenics. A similar, detailed analysis of the WB Vocabulary performance of brain damaged patients by Heilbrun (1958b) revealed no significant differences between such patients and physically ill patients either in terms of accuracy (standard scoring) or mode of response (categorical, de-

scriptive, equivalent, or functional). Thus, the concept of "latent aphasia" was not confirmed. Heilbrun (1958a) also assessed the discriminative effectiveness of D between brain damaged patients, psychotics, neurotics, physically ill, ward attendants, and college students. He concluded that:

> despite the established sensitivity of the D test to cerebral pathology, the test still falls short of being a useful method of discriminating between brain damaged and non-brain damaged.

Measurement of Scatter

Difference scores. Shortly after publication of the WAIS, Jones (1956) and McNemar (1957) cautioned that differences between subtests may not have diagnostic significance since the distribution of difference scores for "normals" extends considerably beyond the point of statistical significance determined by the standard error of measurement, e.g., 30% of even the standardization population received a statistically reliable difference score between certain subtests. The median reliability of these difference scores was reported by McNemar as being .60; hence, much of the difference score variance is attributable to errors of measurement. Fisher (1960), correcting Wolfensberger's calculations (1958), presented a table for determining the significance of a difference between VIQ and PIQ on the WAIS and WB. Field (1960b), like Jones and McNemar, emphasized the distinction between the "abnormality" and the "reliability" of a WAIS difference score and presented useful tables indicating abnormality and statistical reliability of VIQ-PIQ differences and the reliability of subtest discrepancies singly or in combinations.

The abnormality-reliability distinction is easily seen by noting that a VIQ-PIQ discrepancy of approximately 25 points occurred once in every 100 subjects in the standardization population; thus, a greater discrepancy might be considered significant or "abnormal" in a statistical sense (see tables by Fisher & Field). On the other hand, a VIQ-PIQ discrepancy of approximately 13 points would occur only once in 100 times by chance, i.e., because of errors of measurement associated with the IQ scores involved in the comparison. Consequently, a VIQ-PIQ discrepancy of 13 or greater is not likely to be spurious in the sense of a measurement error, but such "real" differences are not unusual in the general population until they reach the magnitude of 25 IQ points or more. Apparently this distinction has not been thoroughly understood or has been disregarded. Even Wechsler (1958, p. 160) said that "in most instances a difference of 15 or more (IQ) points may be interpreted as diagnostically significant" and at a later point that "a deviation of two or more scaled score units on any subtest from the [subtest] mean is a convenient cut-off point" in defining what constitutes an "abnormal deviation." However, according to Field's table involving the *reliability* of differences, a subtest must deviate by at least 5.75 weighted score points from the mean of the remaining subtests in order to be significant at the .05 level.

Griffith and Yamahiro (1958) reported the reliability or stability of subtest scatter in a heterogeneous group of 55 neuropsychiatric patients over an interval of 1–10 years (mean duration 42 months). The rank-order correlation between subtest scores averaged .51 with the higher rho's being associated with test-retest comparisons involving the same form and shorter intervals. They cautiously conclude that:

> whether the patterns of deviation do or do not

have personality or psychodiagnostic validity, the reliability is such that they might have.

Subtest deviation scores from Vocabulary would seem to be a dependable procedure for psychiatric patients since Kasper (1958) found no significant relationship between ratings of "morbidity" (Lorr's Multidimensional scale) and Vocabulary for psychiatric patients.

Intellectual efficiency and potential. Since the inference of intellectual efficiency is sometimes made from a minimum of intratest scatter on the WAIS Vocabulary, Fink and Shontz (1958) analyzed 100 random protocols from Wechsler's standardization and 100 from physically ill patients in order to determine the frequency of 0-, 1-, and 2-point responses for each Vocabulary item. Several deviations from the expected frequency for stimulus words were noted: e.g., WINTER, BREAKFAST, FABRIC, SLICE, ENORMOUS, SENTENCE, REGULATE, and REMORSE all yielded more one-point responses than expected for both groups. Brown and Bryan (1957) concerned themselves with an "altitude quotient" (IQ based upon the two highest subtest scores) as an estimate of intellectual potential in 270 young, "nonclinic" WB subjects. The mean difference between FSIQ and the altitude quotient was 24.6, with a standard deviation of 8.1; this difference tended to diminish with increased intellectual maturity (CA) and higher IQs. A correlation of .87 was found between the IQ and the altitude quotient in this group.

Mahrer and Bernstein (1958) explored performance on repeated Wechsler Verbal subtest administrations. They urged subjects to give as many answers as possible and scored only the best. IQs continued to ascend upon successive administration and they feel that this novel approach gives a good indication of intellectual potential. This method was compared by Thorp and Mahrer (1959) with four other more easily calculated estimates of potential intelligence: (*a*) intersubtest variability; (*b*) prorating the IQ from the highest subtest score; (*c*) prorating the IQ from Vocabulary; and (*d*) prorating the IQ from the three highest subtests weighted by 2.5, 1.5, and 1.0, respectively, from highest to lowest. For 60 neuropsychiatric military patients, only Methods *b* and *d* involving the higher subtests yielded high correlations (.80 to .90) with the potential IQ estimated by the more laborious method. Yet, Mahrer and Bernstein's method yielded a higher estimate of potential intelligence "in almost every case" than the corresponding estimate by the other methods. These investigators also found a negative correlation (−.41) between the FSIQ and the increase in IQ when potential was estimated which seemed largely attributable to IQs over 105, suggesting a ceiling effect.

Scatter and diagnosis. By tallying the incorrect WAIS PC responses of 110 normal females and 110 female psychiatric patients, Wolfson and Weltman (1960) determined the errors characteristic of female psychiatric patients. As one might expect, psychotics were more likely to give a unique response than were neurotics or personality disorders, and 81% of the patients gave at least one unique response. Trehub and Scherer (1958) investigated the individual intersubtest variability within a sample of psychiatric patients composed of 166 (61.7%) schizophrenics and 103 neurotics or character disorders. Their cutting score indicative of schizophrenia yielded 72.1% correct identification, an improvement of 10.4% over the schizophrenic base rate. The proportion of misclassifications could have been further reduced by using

only the extremes of the distribution; however, this necessitates a corresponding reduction in the number of patients about whom diagnostic statements are made.

An obvious limiting factor in the usefulness of any diagnostic sign is that it may differentiate selected diagnostic groups but not be uniquely associated with a single nosological category. For example, Ladd (1959) found that intersubtest variability was also greater in a brain damaged group than in a comparable neurotic group; Diller (1955) reported an inflated "mean range ratio" in delinquents; and Plumeau, Machover, and Puzzo (1960) found a higher total scatter index for alcoholics. Consequently, other indices are needed to distinguish one pathological group from the other and such are the goals of pattern analysis, to be discussed in a subsequent section.

Summary. A necessary distinction has been drawn between the *reliability* of a difference in IQ points or subtest weighted scores and the *frequency* of occurrence of such differences in specified populations. The cautions against confusing the two concepts should be heeded. Measures of intersubtest scatter frequently distinguish groups of delinquents, schizophrenics, and organics from normals; however, the diagnostic value of this "sign" alone is negligible since it is clearly not unique to any one diagnostic group or sufficiently discriminative to be reliable in the individual case. There are several fairly reliable but not necessarily highly correlated methods of estimating intellectual efficiency or potential, but we must wait hopefully for evidence regarding the usefulness of such measures.

Pattern Analysis

The performance of Wechsler's group of adolescent psychopaths (1944) was characterized by PIQ > VIQ, OA+PA > BD+PC, and PA > all other subtests. Using a sample of sex offenders ranging from 14 to 64 years old, Wiens, Matarazzo, and Gavor (1959) found the PIQ-VIQ relationship to be supported, while neither Foster's (1959) adolescent recidivists, Field's (1960a) English recidivists, or Panton's (1960) prisoners support it. Foster did find that OA+PA > BD+PC but that PA exceeded only BD and D. Graham and Kamano (1958) found a pattern similar to Wechsler's psychopathic group in a sample of inmates of a federal institution *only* when they were also classified as unsuccessful readers; the "successful readers" did not yield the predicted pattern. Purcell (1956) found that in his sample of Army trainee delinquents BD was least impaired, and that the most frequent offenders did poorest on C, V, and A.

A thorough analysis of the WB performance of 87 male and 80 female juvenile delinquents matched for age, grade placement, and global IQ was made by Diller (1955). The sexes were judged equally endowed with potential intelligence as indicated by prorating the three highest subtests, and both obtained a higher PIQ than VIQ. In terms of factors previously identified by Jastak, the delinquents were impaired in "verbal development" (V, I, C, S), "motivation" (A, D, DS), and mildly so in the "psychomotor area" (BD, DS, I, PA). The sexes differed in that the males were superior in "reality contact" (C, PA, PC, OA), while the females had more "self control." Two individual subtests showed sex differences —PC and DS—with males doing better on the former and poorer on the latter.

With regard to subjects addicted to alcohol: some chronic alcoholics

showed evidence of pathology (clinically as well as test-wise) typical of the organic as in studies by Kaldegg (1956) and Tumarkin, Wilson, and Snyder (1955); while other alcoholics, even after 10–30 years of intense indulgence, were reported to show no apparent gross intellectual deterioration (Peters, 1956). Bauer and Johnson (1957) found no significant difference on subtest performance between chronic alcoholics as compared with the general run of neurotics or "functional" psychotics. Plumeau et al. (1960) found that A was lower for "unremitted" alcoholics than for either "remitted" alcoholics or controls.

Effects of Organic Brain Damage

Wechsler's patterns. Wechsler's subtest patterning for organicity was not cross-validated by Everett(1956), Fisher (1958), Ladd (1959), Love (1955), Reitan (1959). Wechsler's observation that PIQ<VIQ was found by both Ladd and Love in their heterogeneous organic samples, in a group of organics with nonfrontal lobe lesions by Morrow and Mark (1955), in a group with right hemisphere damage by Klove (1959), in a group demonstrating poor "spatial integration" by Klove and Reitan (1958), and in a group of normal senescents of superior intellectual ability by Norman and Daley (1959). Eisdorfer, Busse, and Cohen (1959) found PIQ<VIQ for an aged group and Morrow and Mark observed this relationship in their organics grouped by foci.

With regard to Wechsler's "Hold-Don't Hold" ratio: Reitan (1959) found some support for this pattern when using a pathological group as compared to Norman and Daley (1959) who did not when using normal senescents. In this ratio it is assumed that C, I, PC, and OA will be resistive to the effect of factors contributing to intellectual deterioration. Reitan's (1956) organics did not do well on C and I, as compared to the organics seen by Howell (1955); Inglis, Shapiro, and Post (1956); Klove and Reitan (1958); and Morrow and Mark (1955). The organic samples of Klove, Klove and Reitan, and of Morrow and Mark, and Norman and Daley's senescents did not do well on PC. None of the organics assessed by Ladd (1959), Morrow and Mark, or Norman and Daley's senescents did well on OA, although Klove's organics did.

In Wechsler's ratio it is also assumed that D, A, BD, and DS will be most affected by factors contributing to intellectual deterioration. In general, this was supported by the findings of Klove and Reitan (1958), Ladd (1959), Love (1955), Norman and Daley (1959), and Reitan (1956). However, neither Heilbrun (1958a), Reitan (1959), or Ladd found that D was significantly lower for their organics, whereas Klove and Reitan, Morrow and Mark (1955), and Tolor (1956, 1958) did. Klove (1959) found that low D and A were characteristic of his sample of patients with left hemisphere damage only. The findings of Heilbrun (1959), Howell (1955), Klove, and Parker (1957) all attest to the significantly poor performance of organics on BD, and Thaler (1956) found that decrements in BD were directly related to aging. This is contrary to the performance of the organics seen by Fisher (1958), and Inglis et al. (1956), or the senescents seen by Norman and Daley. Neither Fisher nor Howell found that their samples of organics demonstrated any unique difficulty on DS; however, the groups seen by Klove, Klove and Reitan, and Morrow and Mark did. Moreover, the data of Loranger and Misiak (1959), Nor-

man and Daley, and Thaler demonstrate that DS performance, as with BD, declined with age. Yet Hall (1956) observed that the organic pattern, DS+BD<I+V, frequently occurred in nonorganic patients.

Hewson ratio. Everett (1956) found no significant relationship between the presence of organicity and the Hewson ratio while McKeever and Gerstein (1958) found that the Hewson ratio classified 75% of a group of schizophrenics as organics. Bryan and Brown (1957) found that the Hewson ratio identified 27% of a nonorganic group as organic, that 38% of a group of adolescents suspected of having CNS involvement on the basis of clinical data were identified as organic, but that 67% of patients with known organic involvement of a "mild" degree, and 96% of patients with a "moderate" to "marked" degree of organic impairment were correctly identified as organic.

Effects of specific organic involvements. Bressler (1956) found that PIQ significantly differentiated aphasics from normals, but not organics with aphasia from those without. Fisher (1958) found that paretics demonstrated selective impairment on subtests and that paresis affects verbal abilities to as great an extent as performance. Klove and Reitan (1958), Milner (1958), and Reitan (1955) found that patients with left hemisphere lesions do poorer on verbal tests as compared to those with right hemisphere lesions, the latter doing poorer on performance tests. Heilbrun (1956) also found lower verbal scores for left hemisphere lesions but failed to find that their performance scores were better than for the right hemisphere group. Bortner and Birch (1960) found left hemiplegics had more difficulty with BD than right hemiplegics but the *N* was small and the task involved only recognition.

Thaler (1956) found that patients with normal and focal EEG tracings perform better on such tests as V, I, BD, and DS as compared to those with mixed or diffuse tracings. However, Morrow and Mark's data (1955) suggest (*a*) no significant difference in the performance of patients with either focal or diffuse cortical lesions; (*b*) patients with frontal lobe lesions showed only slight intellectual impairment, save on DS, while patients with lesions dorsal to the Rolandic fissure demonstrated a tendency toward greater intellectual impairment; and (*c*) patients with left hemisphere damage demonstrated a tendency to loss in VIQ and PIQ, whereas patients with bilateral lesions showed loss in PIQ only.

Summary. Research findings in this section are at best inconsistent and, hence, inconclusive. One study demonstrated a superiority of predictions based on behavioral data as compared to a few a priori test patterns (Gaston, 1959) and another, the difficulty of even some seasoned clinicians to sort test profiles into gross categories of neurosis, schizophrenia, and organicity (Cohen, 1955); and Frank (1956) found the same inability to sort patterns even when the "sorter" is factor analysis. Yet in spite of the continued equivocality of the findings, faith persists in the assumption that a test of cognitive functions *should* be able to reveal more about a person than just his IQ. This faith may not be completely unjustified.

One might ask whether the supportive evidence might not be chance phenomena, whether the persistent inconsistency of the findings from review to review does not strongly suggest the fruitlessness of attempting to make assessment of Wechsler

patterns. Yet the frequent occurrence of positive studies may be regarded as evidence that analysis of patterns can be meaningful and that something other than the tool itself might account for the failure of the research to provide consistent and definitive answers.

One of the methodological shortcomings is the failure to distinguish between a mean diagnostic group profile and modal patterns of homogeneous subjects in a diagnostic group. While there is only one mean group profile for a sample, several groups of the subjects may form clusters of homogeneous symptoms with rather dissimilar modal patterns. Furthermore, the group profile cannot be expected to conform to any of the modal patterns since it is a statistic and no single subject should be expected to correspond to the mean group profile. Only modal patterns are appropriate for diagnostic purposes. Wechsler (1944) fails to identify the nature of his proposed diagnostic patterns. Since only one is given for each diagnostic group it seems likely that he has proposed the relatively useless group profile; at least, this is presumed by most investigators in checking the validity of his proposals. Only a clear understanding of these simple principles can lead to a respectable research approach to diagnostic pattern analysis.

An analysis of the investigations beyond the results per se suggests that much is still to be desired with regard to the designs of the research. For instance, from a purely methodological point of view, one might wonder whether or not clinical facts are being sacrificed for statistical significance. In light of the many variables other than intelligence and psychopathology that tend to affect subtest performance and greatly expand error variance, the arbitrary limits of the .01 or .05 level of confidence might be too high. Yet a pattern that fails to discriminate between groups at these levels of confidence would seem too weak to use clinically with individuals.

One might also be disappointed at the seeming lack of flexibility and/or creativity regarding the form of these experiments. The majority of the studies employed the matched group design using a statistically simple test of an inference (chi square or *t*). Zero-order statistics are seldom suited to the complex analysis or identification of multidimensional patterns. Of the many studies surveyed in this section only two went beyond the single or simple multiple correlational techniques into factor analysis, only six went beyond a *t* or the utilization of *F* as a multiple *t*, and only three studies made use of an analysis of variance design to test interaction effects.

One might also show concern regarding the samples of subjects upon which the conclusions are based. Samples of organics employed have been observed to contain such disparate kinds of pathology as tumors, vascular pathologies, infectious diseases, various kinds of head trauma, epilepsy, and developmental anomolies. Included in a single sampling have been patients with lesions which have been both focal and diffuse, have involved different lobes, have been uni- and bilateral, or have been both cortical as well as subcortical in nature. Similarily, in the research on the "character disorder," the sorts of behavior included in such a grouping might vary from such offenses as delinquency, to burglary and dope peddling, to assault, rape, and arson.

One might note that McKeever and Gerstein (1958) found that measures of organic deterioration varied systematically with age, and Fry

(1956) found that the process of deterioration was not the same in people with limited intellectual capacity as compared to others. Grouping subjects by the criteria involved in the classification of "character disorder," or some synonymous phrase, proves to be no more valid. Lazzari, Ferracuti, and Rizzo (1958) found significant differences in the mean IQ of samples of delinquents just on the basis of crime committed, i.e., fraud *vs.* rape. Wiens, Matarazzo, and Gavor (1959) found that upon more extensive and intensive study, patients initially diagnosed as "character disorder" turned out to be sociopathic personalities, inadequate personality types, mental defectives, adjustment reactions in adolescence, adult situational reactions, depressive reactions, neurotics, schizoid personalities, and even schizophrenics and organics. Therefore, there is reason to assume that such heterogeneous groupings introduce a variety of systematic effects which would detract from the identification of consistent and meaningful patterns associated with the disorder.

In the experiments reviewed herein, investigators have attempted to offset the effect of certain variables by the method of randomization. Yet there is some doubt (Cohen, 1955) that this is an entirely effective procedure in equalizing the influence of such factors as age, IQ, education, etc. It is still not certain whether some of the confusion in the findings might not be attributed to the inadequacy of such a procedure. No investigators actually sought to determine whether the range of age, education, and IQ within the samples made for a significant lack of homogeneity of the groups.

It would appear that systematic research is still necessary to *satisfactorily* establish diagnostic patterns. We wonder if the present interest in pattern analysis of organic brain diseased patients will persist or will it, like the former search for schizophrenic signs, being unrewarded, evaporate. We hope that Reitan's current use of carefully specified types of organic patients for investigation will yield significant patterns and point the way for similar investigation of homogeneous groups of schizophrenics.

General Summary

The WAIS is a much improved instrument when compared with its predecessors. It measures pretty much the same thing that a number of other standardized methods attempt to do. However, comparative studies of the instrument suffer from methodological shortcomings and rely excessively on correlational techniques and insufficiently on comparisons of mean scores.

The test has quickly become domesticated in the various research and clinical settings and has produced some interesting findings reflecting age differences, sex differences, and relationships with an array of different educational, vocational, socioeconomic, and environmental factors. There is, perhaps, a need to attempt to set up such studies in a broader and deeper theoretical framework rather than to continue isolated forays in the flatlands of pure empiricism. Wechsler (1958) has "become increasingly convinced that intelligence is most usefully interpreted as an aspect of the total personality . . . an effect rather than a cause."

Actually the studies on anxiety, impulsiveness, distrust, etc. included in this review are beginnings in the right direction. Inferring other personality variables from intellectual functioning is really an important

avenue to diagnosis. When the concept of diagnosis is thus more broadly conceived, as personality assessment, we need not concur with Meehl (1960) in his pessimistic prognostication.

The additional work on "scatter," profiles, and patterns has not led us on more solid diagnostic ground. The results with the several nosological categories are inconclusive. Severe methodological shortcomings of the investigations prevent the isolation of modal profiles useful for diagnosis. It is perhaps time to face the challenge embodied in Binder's (1956) study of schizophrenia. Is there a differential intellectual impairment? Binder answered the question in the negative by demonstrating an over-all reduction in schizophrenic functioning when assessed with an instrument (SRA tests) which measures relatively independent abilities. Relatively independent factors of mental ability, isolated from the WAIS, might serve as a sounder basis for future diagnostic studies of nosological groupings.

Finally, we again must mention the inadequacy (heterogeneity) of the criterion—schizophrenia, character disorders, etc. We discussed the issue in detail elsewhere (Rabin & King, 1958) and have recommended "The selection of a specific frame of reference in the determination of samples . . . chronicity, or reactive vs. process" as an avenue and approach to more fruitful research.

REFERENCES

ALLEN, R. M., THORNTON, T. E., & STENGER, C. A. The Full-Range Picture Vocabulary Test compared with two short forms of the Wechsler scale. *J. educ. Res.*, 1956, **50**, 133–137.

ARMITAGE, S. G., & PEARL, D. WB changes over time. *J. clin. Psychol.*, 1958, **14**, 22–24.

BALINSKY, B., & SHAW, H. W. The contribution of the WAIS to a management appraisal program. *Personnel Psychol.*, 1956, **9**, 207–209.

BARRY, J. R., FULKERSON, S. C., KUBALA, A. L., & SEAQUIST, M. R. Score equivalence of the WB Intelligence Scales, Forms I and II. *J. clin. Psychol.*, 1956, **12**, 57–60.

BAUER, R. W., & JOHNSON, D. E. The question of deterioration in alcoholism. *J. consult. Psychol.*, 1957, **21**, 296.

BAYLEY, NANCY. Data on the growth of intelligence between 16 and 21 years as measured by the WB scale. *J. genet. Psychol.*, 1957, **90**, 3–15.

BINDER, A. Schizophrenic intellectual impairment: Uniform or differential? *J. abnorm. soc. Psychol.*, 1956, **52**, 11–18.

BLACKBURN, H. L., & BENTON, A. L. Revised administration and scoring of the Digit Span Test. *J. consult. Psychol.*, 1957, **21**, 139–143.

BLOOM, B. L. Ecological factors in the WAIS Picture Completion Test. *J. consult. Psychol.*, 1959, **23**, 375.

BORGATTA, E. F., & CORSINI, R. J. The Quick Word Test and the WAIS. *Psychol Rep.*, 1960, **6**, 201.

BORTNER, M., & BIRCH, H. G. Perceptual and perceptual-motor dissociation in brain-damaged patients. *J. nerv. ment. Dis.*, 1960, **130**, 49–53.

BREIGER, B. The use of the WB Picture Arrangement Subtest as a projective technique. *J. consult. Psychol.*, 1956, **20**, 132.

BRESSLER, MILDRED B. A study of an aspect of concept formation in brain-damaged adults with aphasia. *Dissertation Abstr.*, 1956, **16**, 568.

BRIGGS, P. F. The validity of WAIS performance subtests completed with one hand. *J. clin. Psychol.*, 1960, **16**, 318–320.

BROWN, M., & BRYAN, G. ELIZABETH. The altitude quotient as a measurement of intellectual potential. *J. clin. Psychol.*, 1957, **13**, 137–140.

BRYAN, G. ELIZABETH, & BROWN, M. H. A method for differential diagnosis of brain damage in adolescents. *J. nerv. ment. Dis.*, 1957, **125**, 69–72.

BUROS, O. K. *The fifth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1959.

CALDWELL, M. B., & DAVIS, J. C. A short form of the Wechsler Intelligence Scale Form II for a psychotic population. *J. clin. Psychol.*, 1956, **12**, 402–403.

CALVIN, A. D., KOONS, P. B., JR., BINGHAM, J. L., & FINK, H. H. A further investigation

of the relationship between manifest anxiety and intelligence. *J. consult. Psychol.*, 1955, **19**, 280–282.

CHAGNON, M. *Utilization de l'Echelle d'Intelligence Ottawa-Wechsler*. Ottawa: Editions de l'Universite d'Ottawa, 1955.

CLAYTON, H., & PAYNE, D. Validation of Doppelt's WAIS short form with a clinical population. *J. consult. Psychol.*, 1959, **23**, 467.

COHEN, J. The efficacy of diagnostic pattern analysis with the WB. *J. consult. Psychol.*, 1955, **19**, 303–306.

COHEN, J. A factor-analytically based rationale for the WAIS. *J. consult. Psychol.*, 1957, **21**, 451–457. (a)

COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290. (b)

COLE, D., & WELEBA, L. Comparison data on the WB and the WAIS. *J. clin. Psychol.*, 1956, **12**, 198–199.

COONS, W. H., & PEACOCK, E. P. Inter-examiner reliability of the WAIS with mental hospital patients. *OPA Quart.*, 1959, July.

DANA, R. H. A comparison of four verbal subtests on the WB, Form I, and the WAIS. *J. clin. Psychol.*, 1957, **13**, 70–71. (a)

DANA, R. H. Manifest anxiety, intelligence, and psychopathology. *J. consult. Psychol.*, 1957, **21**, 38–40. (b)

DAVIS, J. C. The scatter pattern of a southern Negro group on the WB Intelligence Scale. *J. clin. Psychol.*, 1957, **13**, 298–300.

DAVIS, P. C. A factor analysis of the WB scale. *Educ. psychol. Measmt.*, 1956, **16**, 127–146.

DILLER, JULIET C. A comparison of the test performance of male and female juvenile delinquents. *J. genet. Psychol.*, 1955, **86**, 217–236.

DOPPELT, J. E. Estimating the FS score on the WAIS from scores or four subtests. *J. consult. Psychol.*, 1956, **20**, 63–66.

DOPPELT, J. E., & WALLACE, W. L. Standardization of the WAIS for older persons. *J. abnorm. soc. Psychol.*, 1955, **51**, 312–330.

DUNNETTE, M. D., & KIRCHNER, W. K. Validation of psychological tests in industry. *Personnel Admin.*, 1958, **21**, 20–27.

EISDORFER, C., BUSSE, E. W., & COHEN, L. D. The WAIS performance of an aged sample: The relationship between verbal and performance IQ's. *J. Geront.*, 1959, **14**, 197–201.

EVERETT, EVALYN G. A comparative study of paretics, hebephrenics, and paranoid schizophrenics on a battery of psychological tests. *Dissertation Abstr.*, 1956, **16**, 1502.

FIELD, J. G. The performance-verbal I.Q. discrepancy in a group of sociopaths. *J. clin. Psychol.*, 1960, **16**, 321–323. (a)

FIELD, J. G. Two types of tables for use with Wechsler's Intelligence Scales. *J. clin. Psychol.*, 1960, **16**, 3–7. (b)

FINK, S. L., & SHONTZ, F. C. Inference of intellectual efficiency from the WAIS Vocabulary Subtest. *J. clin. Psychol.*, 1958, **14**, 409–412.

FISHER, G. C. Selective and differentially accelerated intellectual dysfunction in specific brain damage. *J. clin. Psychol.*, 1958, **14**, 395–398.

FISHER, G. M. A corrected table for determining the significance of the difference between verbal and performance I.Q.'s on the WAIS and the WB. *J. clin. Psychol.*, 1960, **16**, 7–8.

FISHER, G. M., & SHOTWELL, ANN M. An evaluation of Doppelt's abbreviated form of the WAIS for the mentally retarded. *Amer. J. ment. Defic.*, 1959, **64**, 476–481.

FISHER, G. M., SHOTWELL, ANNA, & YORK, DOROTHY. Comparability of the Ammons Full-Range Picture Vocabulary Test with the WAIS in the assessment of intelligence of mental retardates. *Amer. J. ment. Defic.*, 1960, **64**, 995–999.

FOSTER, A. L. A note concerning the intelligence of delinquents. *J. clin. Psychol.*, 1959, **15**, 78–79.

FRANK, G. H. The WB and psychiatric diagnosis: A factor analytic approach. *J. consult. Psychol.*, 1956, **20**, 67–69.

FRY, LOIS M. A predictive measure of work success for high grade mental defectives. *Amer. J. ment. Defic.*, 1956, **61**, 402–408.

GARFIELD, S. L. An appraisal of Object Assembly on the WB and WAIS. *J. clin. Psychol.*, 1960, **16**, 8–9.

GARFIELD, S. L., & AFFLECK, D. C. A study of individuals committed to a state home of the retarded who were later released as not mentally defective. *Amer. J. ment. Defic.*, 1960, **64**, 907–915.

GASTON, C. O. The predictive power of attitudinal and behavioral indices versus formal test scores on the WB test. *Dissertation Abstr.*, 1959, **19**, 1816.

GOODSTEIN, L. D., & FARBER, I. E. On the relation between A-scale scores and DS performance. *J. consult. Psychol.*, 1957, **21**, 152–154.

GOOLISHIAN, H. A., & RAMSAY, ROSE. The WB Form I and the WAIS: A comparison. *J. clin. Psychol.*, 1956, **12**, 147–151.

GRAHAM, E. E., & KAMANO, D. Reading failure as a factor in the WAIS subtest patterns of youthful offenders. *J. clin. Psychol.*, 1958, **14**, 302–305.

GRIFFITH, R. M., & YAMAHIRO, R. S. Reliability-stability of subtest scatter on the WB Intelligence Scales. *J. clin. Psychol.*, 1958, **14**, 317–318.

GRIFFITHS, J. S. The effect of experimentally induced anxiety on certain subtests of the WB. *Dissertation Abstr.*, 1958, **18**, 655–656.

GUERTIN, W. H. Auditory interference with Digit Span performance. *J. clin. Psychol.*, 1959, **15**, 349.

GUERTIN, W. H., FRANK, G. H., & RABIN, A. I. Research with the WB Intelligence Scale: 1950–1955. *Psychol. Bull.*, 1956, **53**, 235–257.

HALL, JULIA C. Two degrees of overt psychiatric disturbance and differences among subtest scores of the WAIS. *Amer. Psychologist*, 1956, **11**, 357. (Abstract)

HALL, JULIA C. Correlation of a modified form of Raven's Progressive Matrices (1938) with the WAIS. *J. consult. Psychol.*, 1957, **21**, 23–26. (a)

HALL, JULIA C. Reliability (internal consistency) of the Wechsler Memory Scale and correlation with the WB Intelligence Scale. *J. consult. Psychol.*, 1957, **21**, 131–135. (b)

HEILBRUN, A. B., JR. Psychological test performance as a function of lateral localization of cerebral lesion. *J. comp. physiol. Psychol.*, 1956, **49**, 10–14.

HEILBRUN, A. B., JR. The Digit Span Test and the prediction of cerebral pathology. *AMA Arch. Neurol. Psychiat.*, 1958, **80**, 228–231. (a)

HEILBRUN, A. B., JR. Vocabulary response as related to lateralization of cerebral lesion: An investigation of "latent aphasia." *J. abnorm. soc. Psychol.*, 1958, **57**, 237–239. (b)

HEILBRUN, A. B., JR. An immediate memory modification of the Block Design task: Relative performances of brain damaged and control subjects. *J. consult. Psychol.*, 1959, **23**, 390–394.

HEILBRUN, A. B., JR. Specificity of immediate memory function associated with cerebral cortex damage. *J. ment. Sci.*, 1960, **106**, 241–245.

HILER, E. W. WB intelligence as a predictor of continuation in psychotherapy. *J. clin. Psychol.*, 1958, **14**, 192–194.

HIMELSTEIN, P. A comment on the use of the abbreviated WAIS with homeless men. *Psychol. Rep.*, 1957, **3**, 440. (a)

HIMELSTEIN, P. A comparison of two methods of estimating full scale IQ from an abbreviated WAIS. *J. consult. Psychol.*, 1957, **21**, 246. (b)

HIMELSTEIN, P. Evaluation of an abbreviated WAIS in a psychiatric population. *J. clin. Psychol.*, 1957, **13**, 68–69. (c)

HOLT, R. R., & LUBORSKY, L. *Selecting psychiatric residents.* New York: Basic Books, 1959.

HOPKINS, A. W. American items on the WAIS. *Ont. Hosp. psychol. Bull.*, 1957, 2–3.

HOWARD, W. A note on McNemar's "On abbreviated WB scales." *J. consult. Psychol.*, 1958, **22**, 414.

HOWARD, W. Validities of WAIS short forms in a psychiatric population. *J. consult. Psychol.*, 1959, **23**, 282.

HOWELL, R. J. Sex differences and educational influences on a mental deterioration scale. *J. Geront.*, 1955, **10**, 190–193.

HOWELL, R. J., EVANS, L., & DOWNING, L. N. A comparison of test scores for the 16–17 year age group of Navaho Indians with standardized norms for the WAIS (Arizona and New Mexico). *J. soc. Psychol.*, 1958, **47**, 355–359.

INGLIS, J., SHAPIRO, M. B., & POST, F. "Memory function" in psychiatric patients over sixty: The role of memory in tests discriminating between "functional" and "organic" groups. *J. ment. Sci.*, 1956, **102**, 589–598.

JONES, H. G. The evaluation of the significance of differences between scaled scores on the WAIS: The perpetuation of a fallacy. *J. consult. Psychol.*, 1956, **20**, 319–320.

KALDEGG, A. Psychological observations in a group of alcoholic patients with analysis of Rorschach, WB, and Bender-Gestalt test results. *Quart. J. Stud. Alcohol*, 1956, **17**, 608–628.

KARSON, S., POOL, K. B., & FREUD, S. L. The effects of scale and practice on WAIS and WB I test scores. *J. consult. Psychol.*, 1957, **21**, 241–245.

KASPER, S. Progressive Matrices (1958) and emotional disturbance. *J. consult. Psychol.*, 1958, **22**, 24.

KERRICK, JEAN S. Some correlates of the Taylor manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75–77.

KLOVE, H. Relationship of differential electroencephalographic patterns to distribution of WB scores. *Neurology*, 1959, **9**, 871–876.

KLOVE, H., & REITAN, R. M. Effect of dysphasia and spatial distortion on WB results. *AMA Arch. Neurol. Psychiat.*, 1958, **80**, 708–713.

LADD, C. E. An analysis of the WAIS performance of brain damaged and neurotic patients. Unpublished master's thesis, University of Iowa, 1959.

LAZZARI, R., FERRACUTI, F., & RIZZO, G. B. Applicazione della Scala di Intelligenza WB

Form I su un gruppo di detenuti Italiani. *Atti Convegno Int. Criminol. Clin.*, 1958, **8**, 449–456.

LEVINE, M. "Not alike" responses in Wechsler's subtest. *J. consult. Psychol.* 1958, **22**, 480.

LEVINE, M., GLASS, H., & MELTZOFF, J. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1957, **21**, 41–45.

LEVINSON, B. M. Use of the abbreviated WAIS with homeless men. *Psychol. Rep.*, 1957, **3**, 287.

LEVINSON, B. M. Cultural pressure and WAIS scatter in a traditional Jewish setting. *J. genet. Psychol.*, 1958, **93**, 277–286.

LEVINSON, B. M. A comparison of the Coloured Progressive Matrices with the WAIS in a normal aged white male population. *J. clin. Psychol.*, 1959, **15**, 288–291.

LIGHT, M. L., & CHAMBERS, W. R. A comparison of the WAIS and WB II with mental defectives. *Amer. J. ment. Defic.*, 1958, **62**, 878–881.

LORANGER, A. W., & MISIAK, H. Critical flicker frequency and some intellectual functions in old age. *J. Geront.*, 1959, **14**, 323–327.

LORANGER, A. W., & MISIAK, H. The performance of aged females on five nonlanguage tests of intelligence functions. *J. clin. Psychol.*, 1960, **16**, 189–191.

LOVE, DEBORAH B. Detecting organic brain damage by means of the WB. Unpublished master's thesis, Howard University, 1955.

MCKEEVER, W. F., & GERSTEIN, A. I. Validity of the Hewson ratios: Investigation of a fundamental methodological consideration. *J. consult. Psychol.*, 1958, **22**, 150.

MCNEMAR, Q. On WAIS difference scores. *J. consult. Psychol.*, 1957, **21**, 239–240.

MCREYNOLDS, P., & WEIDE, M. Psychological measures as used to predict psychiatric improvement and to assess behavioral changes following pre-frontal lobotomy. *J. ment. Sci.*, 1960, **106**, 256–273.

MAHRER, A. R., & BERNSTEIN, L. A proposed method for measuring potential intelligence. *J. clin. Psychol.*, 1958, **14**, 404–409.

MASLING, J. M. The effects of warm and cold interaction on the administration and scoring of an intelligence test. *J. consult. Psychol.*, 1959, **23**, 336–341.

MATARAZZO, J. D., & PHILLIPS, JEANNE S. Digit symbol performance as a function of increasing levels of anxiety. *J. consult. Psychol.*, 1955, **19**, 131–134.

MATARAZZO, RUTH G. The effect of anxiety level upon motor learning, level of aspiration and WB subtest performance. *Dissertation Abstr.*, 1955, **15**, 877–878.

MAYZNER, M. S., JR., SERSEN, E., & TRESSELT, M. E. The Taylor manifest anxiety scale and intelligence. *J. consult. Psychol.*, 1955, **19**, 401–403.

MAXWELL, EILEEN. Validities of abbreviated WAIS scales. *J. consult. Psychol.*, 1957, **21**, 121–126.

MEEHL, P. E. The cognitive activity of the clinician. *Amer. Psychologist*, 1960, **15**, 19–27.

MEER, B., & STEIN, M. I. Measures of intelligence and creativity. *J. Psychol.*, 1955, **39**, 117–126.

MIELE, J. A. Sex differences in intelligence: The relationship of sex to intelligence as measured by the WAIS and the WISC. *Dissertation Abstr.*, 1958, **18**, 2213.

MILNER, BRENDA. Psychological defects produced by temporal lobe excision. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1958, **36**, 244–257.

MONTAGUE, E. K., WILLIAMS, H. L., LUBIN, A., & GEISEKING, C. F. Army tests for assessment of intellectual deficit. *U. S. Armed Forces med. J.*, 1957, **8**, 883–892.

MORROW, R. S., & MARK, J. C. The correlation of intelligence and neurological findings on twenty-two patients autopsied for brain damage. *J. consult. Psychol.*, 1955, **19**, 283–289.

MOYA-DIAZ, LAURA, & MATTE-BLANCO, I. Estudio sobre la correlecion entre el test de Wechsler y el test de Matrices Progressives. *Rev. Psiquiat., Santiago*, 1953–1955, **18–20**, 17–28.

MURPHY, D. B., & LANGSTON, R. D. A short form of the WB and the Army Classification Battery as measures of intelligence. *J. consult. Psychol.*, 1956, **20**, 405.

NEURINGER, C. A statistical comparison of the WB Intelligence Scale, Form I and the WAIS for a college population. Unpublished master's thesis, University of Kansas, 1956.

NICHOLS, R. C. The effect of ego involvement and success experience on intelligence test scores. *J. consult. Psychol.*, 1959, **23**, 92.

NORMAN, R. D., & DALEY, M. F. Senescent changes in intellectual ability among superior older women. *J. Geront.*, 1959, **14**, 457–464.

OLIN, T. D. The use of age-scaled scores in the determination of I.Q. equivalents on the WAIS. *Psychol. Newsltr., NYU*, 1958, **9**, 154–159.

OLIN, T. D., & REZNIKOFF, M. The use of Doppelt's short form of the WAIS with psychiatric patients. *J. consult. Psychol.*, 1957, **21**, 27–28.

PANTON, J. H. Beta-WAIS comparisons and WAIS subtest configurations within a state prison population. *J. clin. Psychol.*, 1960, **16**, 312–317.

PARKER, J. W. The validity of some current tests for organicity. *J. consult. Psychol.*, 1957, **21**, 425–428.

PEARSON, J. S., & AMACHER, PHYLLIS L. Intelligence test results and observations of personality disorder among 3594 unwed mothers in Minnesota. *J. clin. Psychol.*, 1956, **12**, 16–21.

PETERS, G. A., JR. Emotional and intellectual concomitants of advanced chronic alcoholism. *J. consult. Psychol.*, 1956, **20**, 390.

PLANT, W. T., & LYND, CELIA. A validity study and a college freshman norm group for the WAIS. *Personnel guid. J.*, 1959, **37**, 578–580.

PLANT, W. T., & RICHARDSON, H. The I.Q. of the average college student. *J. counsel. Psychol.*, 1958, **5**, 229–231.

PLUMB, G. R., & CHARLES, D. C. Scoring difficulty of Wechsler comprehension responses. *J. educ. Psychol.*, 1955, **46**, 179–183.

PLUMEAU, F., MACHOVER, S., & PUZZO, F. WB performance of remitted and unremitted alcoholics, and their normal controls. *J. consult. Psychol.*, 1960, **24**, 240–242.

PRICE, J. R., & THORNE, G. D. A statistical comparison of the WISC and WB, Form I. *J. consult. Psychol.*, 1955, **19**, 479–482.

PRIESTER, H. J. Ein technischer Hin weis zur Darstellung des Testprofels zum Hamburg-Wechsler Intelligenztest fur Erwachsene. *Z. diagnost. Psychol.*, 1957, **5**, 101–106.

PRIESTER, H. J., & KUKULKA, R. Vergleichsentersuchungen zum HAWIK und Binet-Bobertag und zum HAWIE im Bezung auf die Intelligenzquotienten und die Benutzung dieser Tests als Paralleltests. *Diagnostica*, 1958, **4**, 6–16.

PURCELL, K. A note on Porteus Maze and WB scores as related to antisocial behavior. *J. consult. Psychol.*, 1956, **20**, 361–364.

RABIN, A. I. The use of the WB scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410–422.

RABIN, A. I., & GUERTIN, W. H. Research with the WB test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.

RABIN, A. I., & KING, G. F. Psychological studies. In L. Bellak (Ed.), *Schizophrenia: A review of the syndrome.* New York: Logos, 1958.

RABIN, A. I., KING, G. F., & EHRMANN, J. C. Vocabulary performance of short-term and long term schizophrenics. *J. abnorm. soc. Psychol.*, 1955, **50**, 255–258.

RABINOWITZ, R. Performances of hospitalized psychiatric patients on the Kent Emergency Test and the WB Intelligence Scale. *J. clin. Psychol.*, 1956, **12**, 199–200.

REITAN, R. M. Certain differential effects of left and right cerebral lesions in human adults. *J. comp. physiol. Psychol.*, 1955, **48**, 474–477.

REITAN, R. M. Investigation of relationships between "psychometric" and "biological" intelligence. *J. nerv. ment. Dis.*, 1956, **123**, 536–541.

REITAN, R. M. The comparative effects of brain damage on the Halstead impairment index and the WB scale. *J. clin. Psychol.*, 1959, **15**, 281–285.

ROBERTSON, J. P. S., & BATCHELDOR, K. J. Cultural aspects of the WAIS in relation to British mental patients. *J. ment. Sci.*, 1956, **102**, 612–618.

RUBIN-RABSON, GRACE. Item order and difficulty in four verbal sub-tests of the WB scale. *J. genet. Psychol.*, 1956, **88**, 167–174.

SAUNDERS, D. R. On the dimensionality of the WAIS battery for two groups of normal males. *Psychol. Rep.*, 1959, **5**, 529–541. (a)

SAUNDERS, D. R. An outline of Gittinger's personality theory as applied to the Wechsler: The sub-tests considered separately. (Research Memorandum RM-59-3) Princeton: Educational Testing Service, 1959. (b)

SAUNDERS, D. R. A factor analysis of the I and A items of the WAIS. *Psychol. Rep.*, 1960, **6**, 367–383. (a)

SAUNDERS, D. R. A factor analysis of the PC items of the WAIS. *J. clin. Psychol.*, 1960, **16**, 146–149. (b)

SCARBOROUGH, B. B. Some mental characteristics of southern colored and white venereal disease patients as measured by the WB test. *J. soc. Psychol.*, 1956, **43**, 313–321.

SCHNEYER, S. A short form of the WB scale, Form II: The alcoholic outpatients. *Quart. J. Stud. Alcohol*, 1957, **18**, 382–387.

SIEGMAN, A. W. The effect of manifest anxiety on a concept formation task, a nondirected learning task, and timed and untimed intelligence tests. *J. consult. Psychol.*, 1956, **20**, 176–178.

SINES, L. K. Intelligence test correlates of Shipley-Hartford performance. *J. clin. Psychol.*, 1958, **14**, 399–404.

SINES, L. K., & SIMMONS, H. The Shipley-Hartford scale and the Doppelt short form as estimates of WAIS I.Q. in a state hospital population. *J. clin. Psychol.*, 1959, **15**, 452–453.

SINNETT, KATHLEEN, & MAYMAN, M. The

WAIS as a clinical diagnostic tool: A review. *Bull. Menninger Clin.*, 1960, **24**, 80–84.

SLOAN, W., & NEWMAN, J. R. The development of a WB II short form. *Personnel Psychol.*, 1955, **8**, 347–353.

SPIVACK, G., LEVINE, M., & SPRIGLE, H. Intelligence test performance and the delay function of the ego. *J. consult. Psychol.*, 1959, **23**, 428–431.

STACEY, C. L., & GILL, MARIE R. The relationship between Ravin's Colored Progressive Matrices and two tests of general intelligence for 172 subnormal adult subjects. *J. clin. Psychol.*, 1955, **11**, 86–87.

STERNE, D. M. A note on the use of Doppelt's short form of the WAIS with psychiatric patients. *J. consult. Psychol.*, 1957, **21**, 502.

STERNE, D. M. Use of the Ammons FRPV with the long-term chronically ill. *J. clin. Psychol.*, 1960, **16**, 192–193.

STRONG, P. N., JR. Correlation between the Ohio Literacy Test, grade achieved in school, and WB IQ. *J. clin. Psychol.*, 1959, **15**, 71–72.

SULLIVAN, A. Measurement of intelligence in different environments. *Bull. Maritime Psychol. Ass.*, 1957, **6**, 18–23.

TALLENT, N. Manifest content and interpretive meaning of verbal intelligence test responses. *J. clin. Psychol.*, 1958, **14**, 57–58.

THALER, MARGARET. Relationships among Wechsler, Weigl, Rorschach, EEG findings, and abstract-concrete behavior in a group of normal aged subjects. *J. Geront.*, 1956, **11**, 404–409.

THORP, T. R., & MAHRER, A. R. Predicting potential intelligence. *J. clin. Psychol.*, 1959, **15**, 286–288.

TOLOR, A. A comparison of the Bender-Gestalt test and the Digit Span Test as measures of recall. *J. consult. Psychol.*, 1956, **20**, 305–309.

TOLOR, A. Further studies on the Bender-Gestalt and the Digit Span Test as measures of recall. *J. clin. Psychol.*, 1958, **14**, 14–18.

TREHUB, A., & SCHERER, I. W. WB scatter as an index of schizophrenia. *J. consult. Psychol.*, 1958, **22**, 147–149.

TUMARKIN, B., WILSON, J. D., & SNYDER, G. Cerebral atrophy due to alcoholism in young adults. *U. S. Armed Forces med. J.*, 1955, **6**, 57–74.

URMER, A. H., MORRIS, ANN B., & WENDLAND, L. V. The effect of brain damage on Raven's Progressive Matrices. *J. clin. Psychol.*, 1960, **16**, 182–185.

WECHSLER, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

WECHSLER, D. *Manual for the WAIS.* New York: Psychological Corp., 1955.

WECHSLER, D. *Die Messung der Intelligenz Erwachsener: Textband zum Hamburg-Wechsler-Intelligenstest fur Erwachsene.* (German ed. by Anne von Hardesty und Hans Lauber) Stuttgart: Hans Huber, 1956.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* (4th ed.) Baltimore: Williams & Wilkens, 1958.

WEISGERBER, C. A. A note on Diamond's method of scoring the WB Intelligence Scale for vocational aptitude. *J. clin. Psychol.*, 1955, **11**, 311.

WHITEMAN, M., & JASTAK, J. Absolute scaling of tests for different age groupings of a state-wide sample. *Educ. psychol. Measmt.*, 1957, **17**, 338–346.

WHITMYRE, J. W., & PISHKIN, V. The abbreviated WAIS in a psychiatric populations. *J. clin. Psychol.*, 1958, **14**, 189–191.

WIENER, G. The effect of distrust on some aspects of intelligence test behavior. *J. consult. Psychol.*, 1957, **21**, 127–130.

WIENS, A. N., MATARAZZO, J. D., & GAVOR, K. D. Performance and verbal IQ in a group of sociopaths. *J. clin. Psychol.*, 1959, **15**, 191–193.

WITTENBORN, J. R. The theory and technique of assessment. *Annu. Rev. Psychol.*, 1957, **8**, 331–356.

WOLFENSBERGER, W. P. Construction of a table of the significance of the difference between verbal and performance IQs on the WAIS and the WB. *J. clin. Psychol.*, 1958, **14**, 92.

WOLFSON, W., & WELTMAN, R. E. Implications of specific WAIS Picture Completion errors. *J. clin. Psychol.*, 1960, **16**, 9–11.

ZWART, F. M., & HOUWINK, R. H. Een factoranalyse van de subtests van de WAIS. *Ned. Tijdschr. Psychol.*, 1958, **13**, 415–419.

(Received September 29, 1960)

Psychological Bulletin
1962, Vol. 59, No. 1, 27-56

THE RESEARCH USES OF DOLL PLAY

HARRY LEVIN AND ELINOR WARDWELL[1]
Cornell University *McMaster University*

The history of science offers many examples of potentially useful theories that did not realize their promise until appropriate methods had been devised. Once methodology becomes available, a flood of research often follows, which in turn tries the theory and results in new formulations that in their turn wait on experimental devices. Such progress may falter either because of an absence of theoretical speculation or the lack of methods, and it is futile to assign prior importance to one or the other.

Such interdependence becomes clear when we compare the influence that Binet's scales had on the theories of intellective behavior in children to the relative dearth of systematic research on early personality development, although there is certainly no lack of theories concerning the latter problem. However, standardized methods for appraising personality variables in preliterate children are in short supply.

Reasons for the dearth of methods are not hard to find. Research with preschool children presents certain special problems. The instructions and operations must be simple enough for young children to understand. The subjects must have the physical abilities to perform whatever acts are demanded by the method. Perhaps most important, the tasks must entice and maintain interest against a brief span of attention. Further, for personality research minimal determination of the child's behavior as artifacts of measuring devices is desirable.

Few extant methods meet these criteria. Bellaks' CAT (1950) and the Rorschach (e.g., Ames, Learned, Metraux, & Walker, 1952) are widely and profitably used but present problems. They rely solely on children's language responses and are, in the writers' opinions, too dependent on passive, nonacting out, behavior. Likewise, interviews with children, because they depend on the child's understanding of language may introduce many idiosyncratic factors.

Doll play has offered the promise of range and flexibility in personality research. It is the purpose of this paper to summarize the research uses of doll play and to assay the results of the promise which this method offers.

Doll play started as a clinical device. Anna Freud (1928) attributes its first use to Melanie Klein who employed it as a procedure both for the diagnosis and treatment of disturbed children. However, the concern of this survey is with the *research* rather than the clinical uses of doll play. By "research uses" we shall mean that variables have been measured by this method and related to other variables. The studies reviewed cover the period 1933 to 1960.

What is doll play? There are numerous variations, but essentially the young child is presented with a set of dolls—such as a family—and a setting in which dolls are to operate—such as a home—and told to manipulate the dolls while he tells a story about them. The child has an opportunity here to talk as well as to act. Endless changes can and have been rung on this basic

[1] We are indebted to the participants in the workshop on doll play methods at the 1957 American Psychological Association meetings for many suggestions which appear in this paper. We especially thank Mary Ford, Clara Melville, Judy Rosenblith, and Richard Walters for their comments on the manuscript.

theme: the composition of the dolls, the nature of the setting, the amount and kinds of interaction with the researcher, the directions and structure presented to the child, etc. A host of variables has been scored from the children's protocols. Chief among them are aggression, stereotypy, and prejudice. The method appears to be more useful for some of these variables than for others.

Although doll play was used in research before their work, a strong impetus to doll play as a method in the study of personality development was the work of R. R. and Pauline S. Sears at the Iowa Child Welfare Research Station in the mid-1940's. The original studies, after Bach (1945) indicated the potential value of the method, were methodological and will be discussed below. The serendipitous occurrence of marked sex differences in doll play led to more recent work by these investigators and their co-workers on identification in children, providing one case where a characteristic of a method led to new theory, rather than the converse textbook ideal.

Because of the theoretical dispositions of the early investigators, the most frequent variables measured were derived from behavior theory and were indices of acquired drives in children. Hence, more than any other behavior, fantasy aggression has been measured by this technique, and it was the happy confluence of theory and method that this particular behavior is frequently elicited in doll play.

A host of differences exist among the subjects, equipment, and procedures in the studies which will be reviewed. The following sections are organized to indicate the modal findings or procedure and to sketch the range of variations from the typical occurrence.

Method

Subjects

The usual subjects in doll play research have been preschool children. However, children between the ages of 5 and 10 have been used in many investigations, and in three studies the subjects were up to 13 years old (Honzik, 1951; Levy, 1933; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). Subjects at the extremes of the age range have usually required procedural adaptations. Heinicke (1956), in his study of 2-year-olds, found that children of this age did not use dolls as agents of actions. He felt that he gathered meaningful data about his subjects by putting them in a doll play situation, but in view of the types of variables which yielded results—rate of play, calling for parents, seeking the observer's affection, hostility to dolls and other play objects—it would seem that the findings were incidental to the doll play method. At the upper age levels subjects usually have been instructed to regard the dolls as characters in a play or movie. Only one study has used adult subjects—Rosenzweig and Shakow (1937) compared the constructions of play materials by adult psychotics to those of normal adults, and concluded that their subjects responded favorably to the technique.

Most research with doll play has employed white subjects. Occasionally, in studies of racial identification and prejudice Negroes have been used (Goodman, 1952; Graham, 1955; Radke & Trager, 1950; Stevenson & Stewart, 1958). The method has also been used successfully with American Indian groups (Gewirtz, 1950) and with children in a primitive society (Henry & Henry, 1944).

Both boys and girls have served as subjects. The only indication that there might be sex differences in will-

ingness to play with dolls comes from Finch (1954), who reported that among her subjects, aged 3–8, those from all-boy families refused to participate. However, since her procedure involved doll play in the home as well as in the laboratory, the findings may not be typical.

Equipment

There is no standard material for the construction of dolls—they may be made of plastic, wood, clay, celluloid, rubber, stuffed fabric, or pipe cleaners and cardboard. Their clothing may be nonexistent, simple or elaborate, removable or permanent. However, in the majority of studies reported, the dolls were 1.5″–6″ tall, realistically dressed, and were flexible so that they could be bent to standing or sitting positions. The "standard doll play family," to the extent that it exists, consists of father, mother, boy, girl, and baby. This number can either be reduced or expanded to study particular interactions—e.g., restricted to mother and child (Isch, 1952); to mother, baby, and older brother or sister (Levy, 1933); or expanded to include maid (Bryan, 1940); teacher (Bach, 1945; Melville, 1959); grandparents (Halnan, 1950; Johnson 1952); or additional siblings (Bryan, 1940). Sometimes the subject is given dolls which duplicate his own family (Bremer, 1947; Halnan, 1950; Holway, 1949; Johnson, 1952; Radke, 1946; Ryder, 1954), or he is presented with a large number of dolls of different age-sex categories, and given his choice (Goodman, 1952; Henry & Henry, 1944; Korner, 1949).

The dolls are typically presented in, or in front of, some indoor setting. Most common is the use of a five- or six-room house which has fixed wooden or cardboard walls, but no roof. The house is usually filled with realistic, movable doll furniture which has few manipulatable parts. Sometimes no house is used—instead, the child is given furniture which is either organized into "rooms" or lined up in rows (Bryan, 1940; Finch, 1954; Goodman, 1952; Halnan, 1950; Holway, 1949; Johnson, 1952; Korner 1949; Phillips, 1945; Pintler, 1945; Radke, 1946; Robinson, 1946; Ryder, 1954). Occasionally blocks are available, making it possible for the child to construct walls if he desires them (Bryan, 1940; Pintler, 1945). Settings other than houses have been employed in rare instances—e.g., a complete neighborhood (Meister, 1948), a school room (Bach, 1945; Melville, 1959), or a scale model of a backyard filled with play equipment (Bremer, 1947).

Procedure

Typically, the subject is brought into an experimental room, shown the dolls and other equipment, and told that he may play with them in any way he wishes. Sometimes it is suggested that he make up a story (Bach, 1945, 1946; Bach & Bremer, 1947; Hollenberg, 1949; Johnson, 1951; Krall, 1953; Levin, 1955) but even in these cases the direction of the fantasy is left completely to the child.

The interaction between experimenter and subject is usually controlled to some extent—the experimenter may avoid interaction whenever possible (Bryan, 1940; Honzik, 1951); he may limit the frequency of interaction, usually according to the levels established by Pintler (1945), which will be discussed later; or he may control the situation only in the sense of adopting a constant attitude of noninterfering permissiveness and attentiveness (Bach, 1946; Bach & Bremer, 1947; Holway, 1949; Levin, 1955; Ryder, 1954).

In studies whose primary aim is to compare the results of free doll play

with results of other measures of personality (Ryder, 1954; Simpkins, 1948; Witkin et al., 1954) one session of play may be all that is used. Most experiments, however, provide for the analysis of session-to-session changes, usually with two 20-minute sessions a few days apart. Some studies have used more than two sessions (Bremer, 1947; Heinicke, 1956; Hollenberg & Sperry, 1951; Isch, 1952; Johnson, 1951; Phillips, 1945; Pintler, 1945).

In most cases, the session length is determined beforehand, but even though measures are taken only during the standard time, the subject may be allowed to continue playing as long as he wants (Bremer, 1947). Some workers have not limited the session time, but have recorded all responses until the child lost interest (Goodman, 1946; Ryder, 1954). The latter procedure, of course, makes imperative the use of response proportions instead of response frequencies as measures.

In studying specific variables which would be unlikely to occur with sufficient frequency to give useful results under the free play procedure outlined above, investigators have used a more directive approach in which the setting of the story is specified. The measurements may be mainly in terms of the dolls' actions, as when Levy (1933) records what the subject has the older doll do when it sees its baby sibling at the mother's breast, or doll play may frankly be used as an aid to make it easier to talk to children. In the doll play interview used by Ammons and Ammons (1952), the movement of the dolls is often only an adjunct to enable children to express feelings when they are having difficulty in verbal expression. The same seems to be true of Conn's (1938) study of carsickness, and of Levy's (1940) and Conn's (1940) studies of reactions to the discovery of genital differences.

Studies of prejudice (Goodman, 1946; Radke & Trager, 1950) have confronted subjects with direct choices between white and Negro dolls to reveal their concepts of the status of the racial groups and their preferences for them. In addition, Goodman used a story completion technique, in which the subject decided which doll won in cases of conflict.

The story completion technique is not necessarily restricted to the study of one variable. Since the completion of a prestructured story takes only a short time, a variety of situations can be presented, offering the advantage of overall scores as well as specific ones. Stamp (1954) and Walsh (1956) had their subjects complete a number of stories, including one free story which the child made up himself. Several other studies have used a combination of free play and story completion (Halnan, 1950; Johnson, 1952; Winstel, 1951; Wurtz, 1957).

D. B. Lynn (1955) has developed a Structured Doll Play Test (SDPT) which presents the child with 10 situations in a given order, each with a prescribed arrangement of dolls and furniture. The child completes the story, which in some situations involves a clear-cut choice—e.g., between bottle and cup, crib and bed, mother and father—thus facilitating objective scoring. The SDPT has already been used in investigating age and sex differences (R. Lynn, 1955) and the effects of father absence in Norwegian sailor families (Lynn & Sawrey, 1959), and an extensive program of research using the test is planned (Lynn & Lynn, 1959).

Certainly the effort to get a more standardized procedure to insure comparability among studies is worthwhile. At present, the great variety of materials and procedures which have

been employed make such comparisons of unknown significance. One way to overcome this difficulty is to follow the line suggested above—i.e., to develop a standard procedure and use it throughout an extended research program. However, it is readily apparent that no one method can fit the needs of every research. For example, Wurtz (1957), after trying to use responses to incomplete stories as an index of guilt, concluded that the technique was too highly structured for his purposes. It seems that in addition to standardization an attempt should be made to clarify the effects of variations in equipment and procedure in order to help explain results already obtained, and to offer the prospective research worker information that will allow him to select the best conditions for his purposes.

To a large extent, the worker currently faced with a choice of doll play procedures is offered very little advice. The best way to learn how to do a good doll play study still seems to be to collect the "lore" from someone with experience in the area. Some of this information is given as hints in research reports, but it is scattered and because of its dependence on the specific conditions may not be generally useful. For example, Ammons (1950) found that there were significantly fewer refusals to respond in a doll play interview when simple alternatives were given, when the subject was asked what the doll would do rather than what it would say, when the items were affect-loaded, and when the subject was asked to verbalize the feelings of child, rather than adult, dolls. These kinds of "hints" will be useful to anyone planning to use a doll play interview with a sample similar to Ammons' (boys, aged 2–6), but we cannot say whether they have application to free doll play or to other age-sex groups. Similarly, it would be valuable to be able to predict how much the child will identify the dolls with his own family members, but no attempt has been made to find ways of influencing this variable. Bach (1945) reports that among his nursery school subjects, any insistence by the experimenter that the child identify with the dolls led to resistance by the subject. Within the same approximate age range, Despert (1940) found 14 out of 15 subjects who made at least some specific identification of the dolls with their own families, while Finch (1955) reports little success in getting children to act out parental roles in relation to dolls in the laboratory.

The major attempts to evaluate the effects of equipment and procedure have been made in research under the influence of R. R. Sears. Phillips (1945) found that the only effects of giving the subject highly realistic dolls and furniture rather than having him play with unclothed dolls and "furniture" of simple wooden blocks were increased exploratory behavior and less time spent in organizing the materials. Pintler's (1945) study of the effect of organization of the equipment disclosed that when the furniture and walls of the house were arranged in irregular rows instead of being organized into rooms, children spent more time in organizational behavior. Giving the subject a doll family that duplicates his own has been shown to produce more identification with the dolls than does the use of a standard family (Robinson, 1946). In a study comparing yard and house settings, Bremer (1947) found that the use of a house led to more inappropriate organizational behavior, whereas having the dolls placed in a yard setting with picnic, garage, sandbox, slide, and swing produced more nonstereo-

typed thematic fantasies, more theme changes, and more total aggression.

In the investigation of the effectiveness of the experimenter in maintaining rapport and stimulating the child to elaborate themes in play within the experimental situation (Pintler, 1945) it was found that high interaction between experimenter and child (between 15 and 20 of such stimulating interacts in 5 minutes of play) produced more nonstereotyped fantasies, more theme changes, more aggression, and an earlier onset of aggression play than did a low interaction level (less than 5 interacts in 5 minutes). Studying the effect of the length and number of experimental sessions, Phillips (1945) found no differences between the fantasy material produced in three 20-minute sessions and that in a single hour-long session.

The sex of the experimenter also seems to affect results. Subjects show more aggression in the presence of an experimenter of the same sex (Caron & Gewirtz, 1951).

The above summary represents our total substantive information about the effects of procedural variations. Even our information about those variables that have been investigated is limited, since most of them have been studied in isolation or in combination with only one other manipulated variable. Each is tied to the particular age group on which it was used, and has been tested only with respect to a limited number of dependent variables. For example, perhaps the most widely used reference of those in the above discussion is the work of Pintler (1945). Several studies (Bremer, 1947; Jeffre, 1946; Krall, 1953; Phillips, 1945; Robinson, 1946; Scott, 1954; Sears, 1951; Sears, Pintler, & Sears, 1946; Yarrow, 1948) used "Pintler's high interaction level" or "Pintler's low interaction level" as standards which have been demonstrated to have certain effects. There is no doubt that this work is a valuable contribution; however, it appears that there is much knowledge still to be gained on interaction level before complete understanding of its role is reached. Pintler's study used only preschool children, and there are indications that the levels she used may be less successful with older children. For example, Simpkins (1948), using 5–9 year old subjects, tried to use Pintler's high interaction level, but decided that the fantasy material was being directed by the experimenter too much, so she employed a more nondirective attitude. E. Z. Johnson (1951) found that neither of Pintler's levels was satisfactory for third graders, and ended up using an intermediate level.

The effects of many potentially influential variables have never been studied. Thus far, all the studies that have varied the behavior of the experimenter have revealed differences as a consequence. Since it appears that the experimenter is necessary to encourage verbalization of the subjects, his role becomes crucial. Is there a way to standardize "the doll play experimenter"? Or is it more profitable to partial out his influence on the results? It seems to us that the answer to these questions awaits the demonstration of significant differences—e.g., attributable to the "warmth" of the adult, a dimension suggested by workers in the field to be of importance. Only after such characteristics have been identified objectively can a decision be made as to how best control them.

Reliability and Validity

Consistency of Behavior

In this section it is not our intention to discuss scoring reliability, since this form of reliability, as in all observational procedures, depends on

the explicitness of the definitions of the observation categories and on the adequacy of observers' or coders' training. Nevertheless, it surprised the reviewers how often researchers neglected the common research protocol of reporting observer reliability. In many cases, this simple omission makes the appraisal of results difficult.

Two kinds of information exist about the consistency of a subject's behavior in doll play: the comparison of scores across two or more sessions (analogous to test-retest reliability) and comparisons between early and later portions of a single session (as in split-half reliability). As with so much of doll play data, most information exists for the aggression variable. The session-to-session correlations in either amount or percent aggression varies from .50 to .85 with a median intersession correlation of about .65 (Ammons & Ammons, 1949; Gewirtz, 1950; Levin & Sears, 1956; Sears, 1951; Stamp, 1954; Yarrow, 1948). These correlations when interpreted against a background of varying observer reliabilities and session-to-session changes in the incidence of aggression (see below) indicate quite acceptable reliability. It should be kept in mind that test-retest reliabilities of more highly standardized tests of intelligence are within the same range for children of this age.

Ammons and Ammons (1949) in a structured doll play situation, report corrected split-half reliabilities for aggression of .77 for the first session and .75 for the second.

For other than doll play aggression, Bryan (1940) reports a more holistic appraisal of behavioral consistency in doll play, wherein a graduate student matched protocols for two sessions at better than chance level. The intersession period in Radke's study (1946) was 4–5 weeks. The consistency in specific categories ranged from 29% for dominant themes to 67% for such variables as attitude toward the mother.

There are, unfortunately, too few reports of the consistency of behaviors other than aggression to appraise the reliability of other variables.

Validity

Doll play shares with expressive-projective techniques certain serious problems in the determination of validity. Take aggression as an example. Since aspects of this behavior are disapproved in real life and since doll play presumably reduces these social restraints it may be expected that children high in the inhibition of real aggression may be especially aggressive under make believe circumstances. Were the result a substantial negative correlation between real life and fantasy aggression, the purposes of validity would still be well served. However, in any group of children we may expect variations among children in the amount of aggression anxiety and so there may be no negative or positive relationships between real and fantasy aggression.

The problem is the ubiquitous one of whether doll play behavior is replicative of real life or wish fulfilling in relation to real life. Bach (1945) estimates that more than 75% of children's doll play responses is replicative and the writers' experiences tend to support this contention. This further complicates matters because the validity problem would be more amenable to solution if a child were consistent in one mode or the other whereas he probably varies even within a single session. These problems will be taken up again later.

Against this pessimistic backdrop, we may inspect the validity of doll play behavior against the following

criteria: observation of real life behavior, teacher's ratings, other measuring techniques, and questioning the child.

Observation of real life behavior. Several impressionistic accounts do not agree. Despert (1940) reports that doll play home life had "associated emotional expressions not in all cases in accordance with the observations made on their overt social behavior (family or group)" (p. 25). By contrast, Miller and Baruch (1950) and Henry and Henry (1944) say that various types of aggression and sibling rivalry are congruent between doll play and real life.

In a well worked-out study, Isch (1952) compared behavior in doll play during four sessions with the observations of mother-child interaction in two half-hour sessions. The correlations tended to be low for equivalent categories—around $r = .20$ —but Isch believed that fantasy tended to reproduce real life. For example, when the mothers were highly rejecting and highly aggressive the children represented the mother doll as aggressive. In general, aggression was more severe in fantasy than in real life, e.g., burning a doll in the stove.

Two other studies are relevant. R. R. Sears (1947), relating several studies, reports a complicated relationship between aggression in nursery school and in doll play. Children who were least aggressive in preschool exhibited both extremes in doll play aggression, the determining factor being how severely the subjects were punished at home for aggression. Heinicke (1956) says that there is a generally good correspondence between nursery school behavior and actions in doll play by 2-year-olds, a younger group than is usually employed in doll play research. However, it should be remembered that 2-year-olds do not engage in doll play, in the usual sense.

Teacher's ratings. The relationship between teacher's ratings and doll play behavior is unclear for several reasons. For one, the results themselves are contradictory. For another, where the teacher is not rating actions similar to those manifested in doll play but is providing data for predicting doll play behavior, the findings are usually rationalizable, post hoc, by common sense or by one theoretical scheme or another. In line with the latter point, for example, Bach (1945) reports that children rated as "compliant" by their teacher had, in doll play, more fantasies about school, more stereotyped fantasies, and were less aggressive toward the teacher doll. These, and other findings like them, seem reasonable, but before we accept them as evidence of formal validity, the specifications for the rejection of the hypothesis are necessary. Unfortunately, the state of theory in personality development is not yet able to provide such specifications.

The problem of replication versus wish fulfillment particularly troubles the interpretation of the relationships between aggression in the classroom and in the fantasy situation. One prediction, based on a theory of displacement, is that docile children in the classroom will be aggressive in the fantasy situation, but the prediction must further involve the manner in which the child's real life aggressions are handled. Restrictive classrooms appear, for instance, to depress fantasy aggression (Levin, 1955).

As with so much of the doll play data, the findings of different researchers do not agree. Bach (1945) found that children rated as "normally aggressive" showed *less* thematic aggression than did either of the extreme groups. Isch (1952), at

least during the first three of four doll play sessions, found just the opposite. By the fourth session, subjects rated as "strongly aggressive" showed the most fantasy aggression. Korner (1949) found no relationships between teacher ratings of hostility and the manifestation of hostile actions in doll play. Bach (1945) may have a resolution to this dilemma when he reports that there is a closer correspondence between rated and fantasy behavior for those children who "identified" with a doll—called it "I," or protected it, etc.

Two impressionistic attempts at validation disagree so completely that they do little more than confuse the issue. In Bryan's study (1940) teachers could match complete protocols of doll play with the appropriate children accurately only in one out of 20 attempts. By contrast, Walsh (1956) reports 90% agreement between doll play and teachers' ratings on such variables as freedom of action, freedom and adequacy of emotional expression, and response to environmental stimuli.

Relationship of doll play to other measuring techniques. The several studies of predictive validity give a generally more hopeful account of the validity of various doll play measures. Ryder (1954) reports that behavior in doll play agrees with that in balloon play and blocking; Simpkins (1948) found that when the Ames picture stories and doll play were scored on the same categories the agreement was high, although there were more responses—many of them nonthematic—in doll play than in the story situation. Witkin et al. (1954) in a study different from most using doll play, found that children who exhibited much organization in fantasy play tended to be able to resist field influences in perception as ascertained by tilting-room-tilting-chair tests and by rod and frame tests.

Radke (1946) strikes the only dissident note in this rubric of validity, among the authors who treat predictive validity. She failed to find relationships between doll play and projective picture identifications which she used as part of a large battery of measures on preschool children.

Doll play and direct questioning of children. Many of the same factors which make doll play data difficult to understand also influence the ways in which children answer interview questions, so that the relationships—or lack—between the two must be treated cautiously. The agreement in responses to the two questions, Which parent do you like best? and Which one does the little boy (doll) like best? ranges from 25% to 63%. When the inquiry is phrased as "Which *doll* loves other most often?" the agreement between the answer and nonstereotyped doll play goes up to 68.4% (Graham, 1955). The closer correspondence of the second study is reasonably attributed to the likelihood that the child was reporting about his doll play performance itself rather than about the antecedents of the fantasy.

In summary, the findings on validity are not substantial, if by validity we mean the correspondence between doll play and nonfantasy behavior. On theoretical grounds, strong congruence should not be expected. More definitive tests of validity must take the form of construct validity which in turn waits on clear and unequivocal hypotheses.

Areas of Research

One of the qualities of doll play which has made it attractive to researchers is the flexibility with which it can be adapted to different content

areas. Modifications of equipment and procedure have made it possible to study a great variety of human problems "in miniature." Among the areas of doll play research are the following: constructive-destructive tendencies (Ackerman, 1937), father fantasies of delinquent children (Bach & Bremer, 1947), evaluation of play therapy (Bixler, 1942), car sickness (Conn, 1938), concepts of parental roles (Finch, 1955), sibling rivalry in American (Levy, 1936) and in Pilaga Indian children (Henry & Henry, 1944), aggression and aggression anxiety in accident repeaters (Krall, 1953), hostility in allergic children (Miller & Baruch, 1950), adult schizophrenia (Rosenzweig & Shakow, 1937), reactions to the discovery of genital differences (Conn, 1940; Levy, 1940), self-concepts of underachivers (Walsh, 1956), and achievement and work fantasies of industrious children (Melville, 1959).

In fact, the number of variables that have been educed from doll play is so great that we cannot catalog all of them. Instead, five problems which have been investigated extensively will be discussed below in an attempt to summarize the present state of information about them and to illustrate some measuring problems commonly found in the use of the doll play technique. The areas are aggression, stereotypy, doll preference, father absence, and prejudice.

Aggression

Far more than any other behavior, aggression has been investigated by doll play techniques. The investigator has assurance that at some point in the doll play procedure, a substantial number of children will evidence some aggressive acts. The behavior may be verbal, or acting out the aggression, or a combination. Conceptually, aggression has been defined as any act whose intent is to injure, physically or psychologically, another doll or equipment. Operationally, this common definition presents certain difficulties. First is the inference of intent. This part of the definition is designed to eliminate accidental aggressive acts. Since the child is manipulating dolls and furniture in a small space, he will from time to time knock over a doll or a piece of equipment without apparently meaning to. Investigators often want to ignore such fortuitous acts, and, in fact, it is not difficult to distinguish such accidental acts from "intended" aggression. It seems to us that the problem in operation is far less serious than is the inclusion of intent in the formal definition.

Another category of events not covered by the definition but often scored as aggression is the attribution of motives or traits by the subject to a doll; e.g., the boy is bad, the mommy is mean. One way of handling this contingency is to include the subject as a scorable agent of aggression and to count the above two examples as aggression from the subject to the appropriate doll.

In fact, a major virtue of doll play is the freedom it provides the investigator to design a scoring system that fits his problems. The many specific categories which have been scored under the general aggression rubric are illustrated in the middle column of Table 1. They include total aggression, verbal and physical aggression (often interpreted as indirect and direct), mischief, scolding, tangential, displaced, projected, etc. The latency of the first aggressive act in the session has been studied, and usually interpreted as an index of aggression anxiety. The agents and objects of aggressive acts are popular topics of study. A generalization, though, is that when many subcategories of aggression are scored, the incidence in any one category is

TABLE 1

SUMMARY OF STUDIES ON AGGRESSION

Author	Measures of aggression	Independent variables
Ammons & Ammons 1953	Reactions to aggression Counter-aggression Leaving field Verbal expression Appeal to adult Inhibition of aggressive feelings Outcome of aggression Success Failure Objects of aggression Negro doll White doll	Age of subject
Bach, 1945	*Hostility-aggression (% of nonstereotyped) Agent of aggression *Object of aggression	*Sex of subject *Teacher ratings on aggression and compliance in school *Frustration before doll play
Bach, 1946	*Total aggression (% of total acts) involving father doll *Agent of aggression Object of aggression	*Father separation *Sex of subject Mothers' reports of their descriptions of absent fathers to children
Bach & Bremer, 1947	*Fantasy aggression (frequency) Killing Justification of hostile aggression Defensive rationalization Aggression in response to commands *Father doll as agent or object	*Delinquency of subjects (home for prepsychopathic children)
Bremer, 1947	*Total aggression (frequency) Nonstereotyped aggression Stereotyped aggression Suffered aggression (expressing suffering in pain) Chasing or escaping Justification of aggression Nonthematic aggression *Agents of aggression Objects of aggression	*Doll house setting compared to yard setting *Sex of subject
Caron & Gewirtz, 1951	*Total aggression (% of total acts) *Latency of aggression Projection of aggression Agent of aggression Object of aggression	*Sex of experimenter *Sex of subject *Age of subject
Gewirtz, 1950	*Total aggression (% of total acts) Direct	*Sac and Fox Indians compared to white children *Age of subject

* Involved in relationships significant at .05 or better.

TABLE 1—Continued

Author	Measures of aggression	Independent variables
	Physical injury Indirect Verbal aggression Discipline Discomfort-causing Agent of aggression Object of aggression Projection of aggression Displacement of aggression Response attenuation (ratio of direct to indirect)	*Sex of subject *Session-to-session changes
Gewirtz & Caron, 1954	*Physical injury (% of total acts) Latency of aggression Agent of aggression Object of aggression	*Sex of experimenter *Sex of child Session-to-session changes
Hollenberg & Sperry, 1951 Hollenberg, 1949 Sperry, 1949	*Total aggression (% of total acts) Aggressive mischief, disobedience Verbal discipline & aggression feelings—scold, threaten derogation Physical discipline Physical injury to person Physical injury to equipment States of uncomfortable feeling *Projection of aggression Intensity of aggression	Frustration at home (mother interview) *Punishment for aggression at home (mother interview) *Sex of child *Session-to-session change *Disapproval of aggression during experimental session
Holway, 1949	Total aggression (frequency)	Mother interview data on (a) Strictness of feeding schedule (b) Mothers feeling tone on feeding schedule (c) Number of months breast fed (d) Age begin toilet training
Jeffre, 1946 Isch, 1952	*Total aggression (frequency) *Total aggression (% of total acts) Agent of aggression *Object of aggression	*Teacher ratings on aggression *Session-to-session changes Observed mother-child interaction (rejection, aggression)
Johnson, 1951	*Total aggression (frequency) *Aggression mischief *Verbal aggression *Physical injury to person or equipment } *Contra-social *States of uncomfortable feeling Verbal discipline Physical discipline } *Prosocial *Agent of aggression *Object of aggression	*Age of subject *Sex of subject *Session-to-session changes

TABLE 1—Continued

Author	Measures of aggression	Independent variables
Korner, 1949	Total aggression (frequency) (Listing of types of aggression observed)	Content of incomplete story Hostility ratings based on parent & teacher ratings
Krall, 1953 Krall, 1951	*Total aggression (frequency) Action aggression *Verbal aggression Aggressive anxiety Differences between verbal aggression & action aggression *Latency of aggression Inhibition of aggression Displaced aggression Projected aggression	*Accident prone compared to accident free children *Sex of subject Session-to-session changes
Levin 1953	*Total aggression (% of total acts)	*Sex of subject *Severity of punishment of aggression at home (mother interview) *Session-to-session changes *Quarreling & fighting in class (teacher judgment)
Levin, 1955	*Total aggression (% of total acts)	*Sex of subject *Dominance-control of classroom teacher (observed) Session-to-session changes
Levin & Sears, 1956	*Total aggression (% of total acts)	*Sex of subject *Identification with parent (mother interview) *Sex of usual punisher (mother interview) Severity of punishment for aggression toward parents (mother interview) *Session-to-session changes *Ordinal differences Socioeconomic status
Levin & Turgeon, 1957	*Total aggression (% of total acts)	*Mother's presence at doll play session *Stranger's presence at doll play session *Sex of subject
Levy, 1936	Prevention of hostility Direction of hostility (order of attack on different objects) Forms of hostility Mild Simple assault Primitive hostility Self-punishment & retribution	Sibling rivalry problems of subjects

TABLE 1—Continued

Author	Measures of aggression	Independent variables
Miller & Baruch, 1950	Presence-absence of *Direct hostility *Indirect hostility Displaced hostility *Hostility against self	*Allergic compared to nonallergic problem children
Phillips, 1945	*Total aggression (frequency)	Realism of materials Session length *Session-to-session changes
Pintler, 1945	*Total aggression (frequency) Latency of aggression	*Experimenter-subject interaction *Organization of materials *Session-to-session changes
Pintler, Phillips, & Sears, 1946	*Total aggression (frequency)	*Sex of subject
Robinson, 1946	Total aggression (frequency) Stereotyped Nonstereotyped Agent of aggression *Object of aggression	*Type of doll family: standard or duplicate of subject's family Presence or absence of sibling in subject's family
Ryder, 1954	*Rating of aggressive feeling Rating on inhibition of aggressive feeling Total aggression (frequency)	*Father separation *Sex of subject
Scott, 1954	Total aggression (frequency) *Agent of aggression *Object of aggression	*Separation from parents
Sears, 1951	*Total aggression (frequency) Nonthematic Thematic *Bodily injury *No bodily injury Nonpersonal aggression (by dolls toward nonpersonal objects) Trouble as result of demons, catastrophes, or imaginary characters *Latency of aggression Agent of Aggression *Object of aggression	*Sex of subject *Age of subject *Sibling status *Father absence *Session-to-session changes
Sears & Pintler, 1947	Agent of aggression *Object of aggression Content of aggression	*Sex of subject
Sears, Pintler, & Sears, 1946	*Total aggression (frequency) Agent of aggression *Object of aggression	*Sex of subject *Age of subject *Father separation *Session-to-session changes

TABLE 1—Continued

Author	Measures of aggression	Independent variables
Stamp, 1954 (story completions)	*Direct (% total aggression: self doll→parent) *Indirect (% total aggression: self doll, with implied intention; or →parent but not by self) Directed→self (% total aggression) *Displaced (remaining % total aggression)	*Teacher ratings of subjects as "rebellious" or "submissive" *Sex of subject Session-to-session changes
Yarrow, 1948	Agent of aggression Object of aggression *Total aggressive acts *Nonstereotyped aggression *Stereotyped *Tangential aggression *Latency of aggression	*Sex of subject Experimentally induced frustration, antecedent to doll play (a) Failure (b) Satiation *Session-to-session changes

so small that they are combined for purposes of analysis into large groupings such as "total aggression" or "direct" and "indirect" aggression, etc.

The tendency to proliferate basic categories and to recombine them into various indices presents a difficult problem for comparing and evaluating studies. Since a large number of combinatorial indices are possible from a few basic variables and since experimenters choose for theoretical or other reasons to form different combinations, studies which should be comparable are not. The evaluator is also tempted to think that many of the combinations and arithmetic manipulations of scores were reached post hoc and to wish for replications of findings.

As illustrative of the large amount of research on aggressibn, only a few topics will be discussed in detail: sex and age influences; session-to-session changes; and the child rearing antecedents of total, displaced, and projected aggression.

Age, sex, and aggression. The single best documented finding using the play technique is that boys are more aggressive than girls. Still, in spite of the overwhelming evidence on this point there are a few contradictory or nonconfirmatory results. Krall(1951) reported more aggression among the girls in her sample, but a careful check on her data is best interpreted as no rather than reversed sex differences. Likewise, Henry and Henry (1944) reported no sex differences in aggression for Pilaga Indian children, and Hollenberg and Sperry (1951) found none among Iowa City nursery school subjects. Since the findings are so overwhelmingly in the other direction the burden of explaining the dissenting results must fall on these few investigators.

E. Z. Johnson (1951) adds an important result to the repetitive "boys more than girls" data. She found that boys do exceed girls in physical aggression, but that girls show more verbal aggression than do boys. This finding is reasonable in light of the findings on overt—nonfantasy—aggression.

Johnson's finding in regard to age of the subjects is also provocative. Younger children show more of what she calls "contrasocial" aggression,

while older children's aggression is more "prosocial," usually depictions of the parents punishing the children. As one compares the 5- with the 8-year-olds, the usual sex difference in aggression decreases. Caron and Gewirtz (1951) confirm this finding. P. S. Sears (1951), on the other hand, found that the sexes become more different in this respect as they are older, but it must be remembered that her subjects ranged in age from 3 to 5, which is younger than the youngest group in the other two studies citing age differences in fantasy aggression.

Session-to-session changes in doll play aggression. Second only to the consistent finding that boys are more aggressive than girls, is the ubiquitous result that children are more aggressive in the second compared to the first session of doll play (e.g., Hollenberg & Sperry, 1951; Levin & Sears, 1956; Sears, 1951). Although the amount of aggression increases, children tend to maintain their relative rank order in aggressiveness (Sears, 1951). The above findings apply to the first two sessions. When children participated in more than two sessions, aggression in the later sessions presented a more complicated picture. For example, Jeffre (1946) reported that across four sessions, aggression toward the experimenter and equipment increased, which seems a likely reflection of a child's frustrated boredom with the doll play task. Pintler (1945), using three sessions, found that the latency of the first aggressive act decreased in the later sessions.

The increase in aggression appears to be related to amount of experience in doll play and not particularly to the interval between sessions. Phillips (1945) compared doll play performance in a single one-hour session to three 20-minute sessions. The changes that occurred between the first and final thirds of the massed session were similar to those between the first and last distributed sessions.

A reasonable explanation of this common finding is that the child learns with time that the restraints against the expression of aggression are not operative in doll play and hence he may vent his impulses more freely. The fact that when a stranger is introduced into a second session the usual increase in aggression does not occur lends experimental credence to this interpretation (Levin & Turgeon, 1957).

Child rearing antecedents of doll play aggression. The hypotheses relating certain child rearing practices to aggression in doll play have come from psychoanalytic and behavior theory. The setting of doll play is thought of as a situation relatively free from real life restraints and so appears on a similarity dimension with home and school, but different enough from the real life settings so that the restraints against aggressive expression are less potent. If, therefore, aggression is punished at home, such actions are less likely to occur at the point of punishment but will be manifest in the safety of doll play. The general hypothesis has been that there is positive correlation between severity of punishment at home and the incidence of aggression in doll play. One shortcoming of the displacement hypothesis is that it does not predict a higher frequency of incidence in doll play than the less severely punished condition—only that such denied behaviors will appear in fantasy but not in real life. A conflict drive hypothesis has been added to the original displacement one to cover this lack (Whiting & Child, 1953, p. 353). This additional hypothesis postulates a drive increment due to the subject's desire to express aggression and his fear of such expression. Since drive operates multiplicatively, the combination of

hypotheses covers the prediction of severe parental punishment leading to high fantasy aggression.

What is the evidence for this hypothesis? Hollenberg and Sperry (1951) reported confirmation. In an earlier summary report of research performed under his direction, R. R. Sears (1947) found the predicted state of affairs: those children who were most severely punished at home were most aggressive in doll play.

An attempt to replicate this finding further entailed fairly elaborate changes in the hypothesis (Levin & Sears, 1956). On a larger and more varied sample—the previous studies were done with university nursery school groups—the simple "punishment leading to aggression" hypothesis did not hold. Rather, doll play aggression was shown to be predictable from a combination of the sex of child, the real life agent of punishment, and the nature of the child's identification with his parent, as well as the severity of punishment. In general, these findings lend themselves more easily to a replicative rather than to a displacement interpretation. Taken in the light of E. Z. Johnson's (1951) finding that older children evidenced more prosocial aggression, the more mature, identified children may be portraying the parental punishment that they have experienced. It is interesting to note that real life aggression among primary school children was predictable from much the same variables as the fantasy behavior in the Levin and Sears' study (Eron, 1958).

One other study based on the displacement hypotheses obtained completely contradictory results (Levin & Turgeon, 1957). The prediction that the presence of the mother at the doll play session would redintegrate aspects of the home and reduce the freedom of doll play was not borne out. The opposite finding emerged; aggression was more frequent before the mother compared to the control condition. The investigators called the original hypothesis into question and suggested that there are characteristics of the doll play situation which make doubtful its use as a point on a simple freedom-from-inhibition dimension.

Wurtz (1960) in a recent theoretical statement argued that mild aggression anxiety should facilitate attenuated aggression in doll play. He found some confirmation for this notion in a reanalysis of earlier data reported by P. S. Sears (1951) and Sears, Pintler, and Sears (1946), when the index of attenuation is based on the use of child compared to adult dolls as agents and objects of aggression.

In addition to thinking of total doll play aggression as a manifestation of displacement, the same phenomenon has been studied within the doll play situation itself. If a child has been punished for aggression, the depiction of this punishment in doll play should arouse more anxiety than in cases of aggression toward a doll less similar to the performer. This conceptualization creates substantial difficulties. It implies that although doll play in general is not very inhibiting, there is still sufficient anxiety to influence the choice of dolls that act as the objects of hostility. We might expect, therefore, a mild and not very consistent effect on the choice of objects of aggression. The unreliability should be compounded by the low incidence of acts which determine any displacement score. For example, if 15% of all doll play units are aggressive and this percentage is divided among five dolls equally, we are dealing with expected displacement scores of 3% of the total number of acts, and the unreliability of this miniscule proportion is obvious.

The implications of the displacement hypothesis for understanding pro- and contrasocial aggression are especially difficult to justify. If severely punished, highly identified children accurately replicate their parents' punishment in doll play, they would be showing little displacement although they have experienced severe punishment.

The final comment on doll play analysis of displacement is, to our thinking, most serious and applies equally below to the discussion of projection. How are the doll agents or objects of aggression to be ordered for the analysis of the two defense mechanisms? Most often, the assumption is made that the child uses the doll most similar to himself as the point of origin on the similarity continuum. For children who are strongly identified with their parents, this assumption is suspect. Granted this point, however, the additional points create greater difficulties. Should the grouping be by sex or age? Does the dimension for a girl go: girl (G), mother (M), boy (B), father (F), baby (bb); or G, B, M, F, bb; or, perhaps, G, M, F, B, bb; etc.? All of these are empirically answerable, albeit difficult, questions. One possibility is that the dimension is an idiosyncratic, response mediated one. Another tack may be that the nature of the dimension varies depending on the behavior being studied; i.e., one sequence of dolls for aggression, another for dependency, etc. As will be pointed out in the final section of this paper, there is little evidence that can be brought to bear on these questions.

An analog to displacement within the doll play session is "projection," which is defined in terms of the doll agents of aggression. Presumably, a doll most similar to the child carrying out hostile acts represents projection of the child's hostile impulses to the doll. The above comments on displacement also apply to this mechanism.

A number of studies relate the agents and objects of aggression to demographic characteristics of the child, as can be seen in Table 1. However, a direct test of the displacement or projection formulations requires information about the nature of the child's aggression anxiety as well as the dolls he chooses to initiate and receive hostile acts. Only one study yields this information directly (Hollenberg, 1949). She found that children who were severely punished for aggression at home projected aggression more in doll play than did less severely punished children. Comparable data on displacement are not available.

In summary, the demographic and practice correlates of doll play aggression are clear and substantial. However, the problems of greater theoretical interest—the child rearing correlates of doll play aggression—must, because of their conceptual unclarities and inconsistent results, remain open questions. A thorough test of the displacement model would require information about the anxiety attached to the expression of aggression at home, the amount of such behavior actually exhibited at home, the instigation to aggression, and the amount of aggression shown in doll play. A questionable assumption is that the instigation to aggression is more or less the same in doll play as in the home—that it is a characteristic of the person independent of the situation. No single study fulfills more than one or two of these requisites.

Stereotypy

Many doll play studies have categorized routine, habitual actions, "doll action appropriate to the time, place, situation, and characters involved" (Phillips, 1945). These be-

haviors are most often termed "stereotyped," although they sometimes have been labeled "realistic" or "routine role." Such doll actions usually constitute a considerable part of the total acts in a session. Krall (1953) reports that stereotyped thematic responses constitute 45% of all responses made in doll play, and Bach (1945) and E. Z. Johnson (1951) report that 59% and 66% of thematic responses are stereotyped actions.

The most consistent finding with regard to stereotypy is a sex difference: girls show more stereotyped behavior than do boys (Bach, 1945, 1946; Bremer, 1947; Honzik, 1951; Pintler, Phillips, & Sears, 1946; Yarrow, 1948). This finding might be attributable to the greater familiarity of girls with doll playing, but since it occurs from age 3 onward, it would seem more likely to be related to greater inventiveness of young boys. As a case in point, Tuddenham (1952) reports that first, third, and fifth graders recognize that the "typical girl" is less daring than the "typical boy."

The amount of stereotypy decreases from session to session (Bach, 1945; Phillips, 1945; Yarrow, 1948), a fact which may be explained in several ways. The higher incidence of aggression in later doll play sessions may displace stereotyped responses. Also, the relaxation of restraints in the second session which yields more aggression may also lead to more nonstereotyped, nonaggressive behaviors. It seems natural that a child faced with a new situation would first represent the most highly practiced behaviors—the routine acts of the home.

Attempts have been made to relate stereotypy to adjustment. Bach (1945) reported that children whom teachers rated as being "well adjusted" showed a higher rate of decrease of stereotypy over sessions than did "poorly adjusted" children. Holway's (1949) findings on "realistic" play, which seems to be closely related if not identical to stereotyped play, show that at the end of therapy, children play more realistically using less fantasy, aggression, or tangential (nondoll) play. Holway's study attempted to relate doll play to child rearing variables. She found that realistic play was positively related to the amount of early self-regulation in feeding and the number of months the child was breast fed.

In Holway's (1949) sample of 3–5 year olds, there was no correlation between realistic play and either CA or IQ. However, Graham (1952), comparing seven "bright" primary school children with seven "dull" ones, found that the brighter children used more stereotyped responses.

Aside from the sex differences, session-to-session changes, and possible IQ differences, there have been no other substantial findings with regard to stereotyped play. In studies of delinquents (Bach & Bremer, 1947), accident repeaters (Krall, 1953), and various methodological explorations reported above (Phillips, 1945; Pintler, 1945; Robinson, 1946), no significant differences in the amount of stereotyped behavior were found between experimental and control groups. In the area of parent separation, the results are not consistent—Bach (1946) found that father-separated children showed more stereotyped fantasies about home life, whereas Scott (1954) reported that institutionalized children indulged in less stereotyped play than did children living with their parents.

The stereotype category is usually regarded as a residual category rather than as a major interest. A recent study indicates that it may have some predictive value if further anal-

yzed. Melville (1959) found that children who spend a large proportion of their school time working industriously use the "work routine" category (that portion of stereotyped behavior which is work oriented) in fantasy more than do less industrious children. Note that this is a more or less direct replication in doll play of observed real life behavior. Melville's study suggests that a finer breakdown of the stereotypy category might be profitable.

Doll Preference

Although many doll play studies record which dolls were used as agents and objects of fantasy acts, few of them report analysis of doll usage in any detail. The greatest interest in this variable has been evidenced by researchers in the areas of aggression and the effects of separation from parents, and the results are presented in the appropriate parts of this paper.

Probably the best substantiated generalization to be made about this topic is that subjects tend to prefer the same sex parent doll to the parent doll of the opposite sex. This tendency shows some increase with age. The finding has not appeared in every study—e.g., Graham's (1952) subjects, regardless of sex, tended to use the mother doll more than the father doll—but significantly greater use of the opposite sex parent has not been reported. E. Z. Johnson (1951) found that while in portrayal of routine (stereotyped) behavior all subjects used the mother more often than the father doll, the greatest session-to-session increase in the use of the father occurred among older boys. In a nursery school sample (Sears et al., 1953), the girls used the mother doll more than the father, while the boys used the two dolls equally, thereby employing the father doll more than the girls did.

Five studies which have used relatively structured situations to lead the child to make a direct choice also report same sex preference. Ammons and Ammons (1949) found a father preference among 3- and 4-year-old boys, and a mother preference among 4- and 5-year-old girls, and R. Lynn's (1955) 6-year-old subjects showed a greater preference for the same sex parent doll than did her 4-year-old subjects. Emmerich (1959) had his subjects complete stories using first the adult and then the child dolls. Correspondence between the two sets of behaviors was taken to indicate high identification. He found that preschool children—especially boys—tended to identify more with parents of the same than with parents of the opposite sex. Similarly, highly sex-typed boys depict more nurturance, punishment, and power via the father than via the mother doll (Mussen & Distler, 1959). To get at sex role identification, Rabban (1950) asked children aged 3–9 to select the doll that "looks most like you." Starting at the age of 4, the choices were correct as to sex.

Preschool children who have been reared permissively emphasize the adult dolls in their fantasy productions (Levin, 1958). This finding may be interpreted in several ways: permissive parents interact more with their children and thereby provide a more frequent adult model, parents who rear their children permissively permit them to explore and practice adult-like behaviors more than do nonpermissive parents, and permissiveness is one of the antecedents of identification with parents which is reflected in the child's preoccupation with adult actions in doll play.

Effects of Separation from Parents

Interest in this area grew out of the problems of wartime father separation, and the majority of studies

have been concerned with the absence of the father rather than the mother from the home. The studies of father absence can conveniently be divided into two groups: those concerned with children currently separated from their fathers, and those of children whose fathers had been absent during the first year or two of the child's life but were living with the family at the time of the study. Bach (1946) studied children aged 6–10 whose fathers were in the service abroad and had been away for 1–3 years. He found that father-separated children, compared to a control group whose fathers were at home, produced fewer doll actions that involved the father doll; enacted a more stereotyped view of family life; and made the father doll more aggressive, less authoritarian, and more affectionate, than did the control group. Using a smaller group of subjects, he found that where the mother described the absent father to the children in deprecatory terms, the children portrayed the father as being more aggressive to his doll children, but as receiving more affection from them; i.e., unfavorable typing of the absent father seemed to produce ambivalent feelings in the children. Another study (Sears et al., 1946; Sears, 1951) found that nursery school children whose fathers were absent from the home did not show the session-to-session increase in aggression that is usually found. In addition, boys (but not girls) whose fathers were absent were less aggressive in their fantasies (Sears et al., 1946). The father-present control group of boys showed most aggression in doll play toward the father doll and the boy doll (sex category), while the boys without fathers showed most aggression toward the father and mother dolls (age category) (Sears, 1951).

Lynn and Sawrey (1959), using the Structured Doll Play Test, have investigated absence of fathers in children of Norwegian sailor families. They found that girls (but not boys) whose fathers were gone were more dependent than the control children. However, on a measure of "maturity" (choice of sleeping in a crib or bed), boys without fathers were less mature than boys whose fathers were at home. In contrast to other studies of father absence, this one also investigated the child's relationships with mother, and concluded by doll play and other techniques that the mothers of father-absent children were more overprotective than were control group mothers.

Studies of homes where the father is currently absent do, then, find substantial results. Positive results have not been so easy to find in studies in the second group—those in which a previously absent father is present in the home at the time of the investigation. Halnan (1950), L. C. Johnson (1952), and Ryder (1954) performed doll play studies as part of the Stanford University research on father relations of war-born children. Only one difference was found between responses of control groups and those of children aged 4–7 who had been separated from their father during the first 2 years of life. In Ryder's study the doll play of the previously father-separated children was rated as revealing more aggressive feeling. Since this was an inferred measure rated by the experimenter and an observer, and since measures of overt aggression in doll play did not show any significant differences in this or either of the other two studies, it must be concluded that there is little evidence of marked effects on the doll play of children temporarily separated from their fathers in early life.

In view of the recent great interest in the effects on the child of separa-

tion from his mother, it is surprising that doll play has not been used to investigate this area. So far as is known, there has been no study using this technique with children living in households where the mother is absent. However, there are two investigations of children separated from both parents. Heinicke (1956) studied 2-year-olds living in residential nurseries because their parents were on vacation, sick, or having another child. He found results which agreed with observations of the subjects in their nursery life—e.g., they sought the affection of adults by crying—but, as has been observed before, most of his results were not specifically concerned with doll play responses. Scott (1954) studied children separated from their parents because they had been institutionalized because of neglect, mental illness in the family, etc. He found that the subjects showed a much greater than average tendency toward "metamorphosis," i.e., the subject himself acted as an authority figure and treated all the dolls as children. It is doubtful that this result should be attributed to parent separation as such; it seems just as reasonable to relate it to the effects of institutionalization.

Reactions to Racial and Religious Differences

Of the studies of children's reactions to Negro-white differences, several (Goodman, 1952; Graham, 1955; Radke & Trager, 1950; Stevenson & Stewart, 1958) have used both Negro and white subjects, while one used only white subjects (Ammons, 1950; Ammons & Ammons, 1953), and another (Clark & Clark, 1947) used Negro subjects exclusively. In this area unstructured doll play, compared to structured, has not produced meaningful results. Graham (1955) recorded the free play of Negro and white subjects with both Negro and white dolls, but made only intraracial analyses of his data. There were no outstanding differences between the two groups. Goodman (1946) used only 24 subjects in the part of her study involving free play, and found no statistically significant differences between Negroes and whites. However, she did uncover several trends which seem worthy of follow-up with a larger group—e.g., Negro subjects tended to assign main roles to white dolls, and seldom revealed positive evaluations of Negro dolls. In her later studies (Goodman, 1952), where no statistical evaluations were made, she reported that doll play was a successful technique, but it is difficult to tell how much success is attributable to the free play method itself since it was used mainly as the introduction to a doll play interview.

It has been much more common, and apparently more profitable, to use controlled methods of exploration like direct questioning about preference for dolls of different color (Clark & Clark, 1947; Goodman, 1952; Radke & Trager, 1950; Stevenson & Stewart, 1958), identification of race (Ammons, 1950; Ammons & Ammons 1953; Clark & Clark, 1947), requiring the child to pair dolls which "go together" (Goodman, 1952), pairing dolls with middle class or slum houses, and with dress-up or work clothes (Radke & Trager, 1950), and various incompleted stories which offer an opportunity for a doll of one color to "win" over a doll of another color (Ammons, 1950; Ammons & Ammons, 1953; Goodman, 1952).

Results obtained from these techniques, sometimes used in connection with more extensive interviewing (Ammons, 1950), are in fair agreement with one another. Negro and white nursery school children appear to be well aware of racial physical

differences (Ammons, 1950; Clark & Clark, 1947; Goodman, 1952). Both racial groups are likely to identify with the white doll when asked "Which looks most like you?" (Clark & Clark, 1947; Goodman, 1946), although with increasing age there is more correct identification until at age 7 a slight majority of Negroes identify with the Negro doll (Stevenson & Stewart, 1958). In addition, some of the Negroes show either confusion or wish fulfillment by insisting that although they are now dark skinned, they had white skins as babies (Goodman, 1946). There have been consistent reports that white dolls are preferred esthetically by white children, while Negro children do not show a clearcut preference for Negro dolls, but instead may either choose the white doll (Clark & Clark, 1947; Goodman, 1946), show only a slight preference for the Negro doll (Radke & Trager, 1950), or show reluctance to make any choice (Goodman, 1952). Interpretation of this result is not unequivocal, since it may reflect past experience with dolls and story book characters who are more often white than colored. More important would seem to be Radke and Trager's (1950) finding that, even when subjects are equated for social class, 5–7 year olds of both races accept the idea that Negroes belong in poorer housing.

While it seems to have been demonstrated clearly by the method of doll play that children are capable of making discriminations on the basis of color, it has not been shown that these discriminations are reflected in fantasy behavior in any consistent way. Ammons (1950) reported that white boys showed a tendency with increasing age to use Negro dolls as scapegoats. On the other hand, analysis of the same data did not reveal any differences in the Negro vs. white dolls in conflict—whichever doll the subject was using at the moment tended to be successful in aggression (Ammons & Ammons, 1953). Stevenson and Stewart's (1958) Southern Negro subjects chose the white doll as the one with whom they would like to play, except at the oldest age level—7 years—where a small majority chose the Negro doll. Goodman (1946) found no social acceptability differences among subjects who had chosen the white doll esthetically. The white doll might be "prettier," but the Negro doll was just as acceptable as a birthday party guest. In further studies, Goodman's (1952) subjects mixed the races indiscriminately in free doll play.

This lack of consistent discriminatory behavior in doll play is paralleled by a similar unconcern in observed behavior. Goodman (1946) found no consistent prejudice in nursery school behavior of her mixed racial group. The results of doll play are congruent with those of other methods in finding a poor correspondence between beliefs and the development of interracial behavior. The evidence on this point has been reviewed by Harding, Kutner, Prochansky, and Chein (1954).

Hartley and Schwartz (1951) described materials and procedures for studying attitudes toward religious groups. Subjects were given three doll families, each of which stands in front of a montage background of photographs, one suggesting a Jewish religious context, one Catholic, and the other a middle class home without any religious symbols. The investigator notices what spontaneous identification the subject makes of the backgrounds, and uses these as a lead-in to a doll play interview with the child. The only data available from the use of this technique are some protocols, but it appears to be easily adaptable to the analysis of group differences.

Experimental Manipulations

It is obvious, at this point, that the bulk of the studies has employed doll play to measure naturally existing characteristics of the subjects, with no attempt to influence these characteristics. By contrast, projective studies of adults have recently used experimental variations both to test specific hypotheses for which manipulation is relevant as well as to ascertain the validity of the measurement (e.g., McClelland, Atkinson, Clark, & Lowell, 1953).

The four experimental studies of doll play divide into two groups: either some experiences of the child prior to doll play or experiences during the course of the procedure are varied. In the first category, Bach (1945), testing a frustration-aggression hypothesis, subjected some of his preschool subjects to a longer rest period than others just before a doll play session. Since a long rest was presumably frustrating, these children elaborated the rest theme in their fantasy output more often and more aggressively than did the short rest group. Yarrow's (1948) results are less clear. He had one group of subjects play with a difficult tinker toy before doll play and compared them to a group who were given an easy task. The frustrated subjects tended to show increased aggression, more tangential play, and distorted thematic play than the other subjects but the results were not statistically significant. When the children experienced antecedent satiation—putting pegs into boards until they refused to continue—they gave more inappropriate thematic units—e.g., sleeping in the kitchen.

To test the effects of aggression anxiety on fantasy aggression, Sperry (1949) compared three groups of children, each of whom participated in four sessions. For one group the experimenter disapproved of the subject's aggression in the second session. The experimenter disapproved of the subject's aggressive acts in the second and third sessions for another group and expressed no disapproval to the control group. Only the group punished in the second session decreased their disapproved acts in the third period (Hollenberg & Sperry, 1951).

Working also with a model of aggression inhibition, Levin and Turgeon (1957) compared two groups of subjects. The first group's second doll play session was observed by their mothers; in the other group a strange adult female was present. Mothers facilitated the children's aggression whereas the stranger inhibited socially disapproved acts.

In general, doll play has suffered from a dearth of experimental treatment. Some experimental operations relevant to the variables being measured would add to the validity of the method and, to judge from other projective techniques, would provide more discriminating measures of individual differences.

Discussion

What can we say now about the doll play technique, which two decades ago appeared so promising? Certainly an overall body of sensible, interrelated findings is not apparent. Where doll play was used in a connected group of studies from one laboratory, coherent results do appear. Otherwise, single investigators performing one or two studies using the method occasionally report interesting results but there are almost as many islands of findings as there are investigators. One might hope that the common method would provide the links between studies, but the flexibility of doll play, both in procedure and scoring of variables, makes the connections among findings tenuous.

In the area of aggression there are results that have been replicated. However, their very redundancy

makes them appear trivial in comparison to what might have been discovered in the years of effort. We may take as fact that young boys are more aggressive than young girls and that children are more aggressive in the second than in the first doll play session. Most other doll play findings have to be hedged with boundary conditions, and restrictions must be put on general statements.

To understand this state of affairs it may be useful to review the virtues and shortcomings of doll play. We believe that the meager payoff comes not from the technique itself, but from the assumptions which underlie the method. First, what should any method of assessing personality provide? Objectivity has not often been a problem in doll play so long as the variables are carefully defined and the scorers are well trained. Reliability has been looked at both in terms of the consistency of behavior and in terms of categorizing agreement among scorers. Besides, the method is not heavily dependent on verbalization, which recommends it for use with young children, and it is interesting to them. The major difficulties appear in understanding what the method is measuring.

Replication and Wish Fulfillment

The basic question that has influenced the understanding of doll play is whether the child is telling about events and hopes and plans which are available to him in his day-to-day world or whether his acts in this setting are otherwise unavailable. The criterion for identifying wish fulfilling fantasies is that nonfantasy expression of the behaviors is prohibited and they are then expressed in fantasy. The prohibitions may be actually imposed on the child or may result from natural conditions: e.g., his color or sex or size. Therefore, the specifications for wish fulfilling fantasies are four: evidence that there are in "real life" some restraints against the expression of the behavior in question, a desire for such expression, little overt manifestation of the behavior, and the appearance of the behavior in fantasy.

Few research studies include all of the requirements of the wish fulfillment model. The studies of parental punishment and fantasy aggression make certain assumptions about the model, but both the results and the assumptions must be questioned since no study clearly replicates another. For example, it is assumed that severe parental discipline inhibits overt expression, yet there is some evidence that punishment and overt aggression are positively correlated (Sears, Maccoby, & Levin, 1957).

In the studies of racial identification and prejudice, the assumptions although not usually specified, are often reasonable. For example, a substantial number of Negro children indicate in doll play that they want to play with white children (e.g., Clark & Clark, 1947). To take this as wish fulfilling behavior we need to know that such interracial play is not possible and actually does not occur. These inferences may be based on sociological characteristics of the child's neighborhood, although it is preferable to test these assumptions directly.

The inclusion of "wishes" under the replication rubric requires some explanation. If the child's wishes are not denied real expression, this category of behavior does not fit our wish fulfillment model. One way of thinking about doll play behavior is that it gives the child an opportunity to express his current experiences and preoccupations. The correspondence between real life and fantasy need not be uninteresting for research purposes. In this type of fantasy the child may give the researcher a picture of his thoughts and actions which would be much more difficult to elicit in an in-

terview. Also, so far as the child's functioning is concerned, replicative fantasy may well provide him an opportunity to practice and develop skills which are transferable to his nonfantasy life.

To take advantage of the wish fulfillment-replication distinction in research, it would be most helpful if a child consistently acted either one or the other type of fantasy. Unfortunately, such is probably not the case. A child may change his emphasis from session to session or may vary the proportions of fantasy within a session. The ideal condition would allow the researcher to categorize a sequence of doll play as wish fulfillment or replicative. Our current knowledge about children's fantasies preclude any such simple procedure although, as we suggest below, there may be some guides in making this decision within doll play itself.

Some researchable problems which would aid in distinguishing and making use of the differences between replication and wish fulfillment are suggested below:

1. Without exception in the doll play studies reviewed the fantasies have been categorized in terms of simple counts of units. Molar sequences of behavior and units of interaction which are now common in observations of adult interaction have not been applied to children's fantasies. For example, if the sequence is "the father spanks the boy and then the boy hits the father" we might be more justified in tolerating the notion for future tests that this is a wish fulfilling episode compared to the "father spanks the boy and the boy cries."

Likewise, doll play actions that indicate that inhibitions are being overcome may be discernible. The two indices that have been used are latency of the first aggressive act and the occurrence of tangential behavior. The latter may be promising if analyzed in a sophisticated fashion. Tangential actions such as looking out the window or engaging the experimenter in conversation which appear irrelevant to doll play may indicate a variety of states. The child may be bored, or unable to think of more actions to portray, or he may indeed be experiencing anxiety over some impulse which is at the threshold of experience. These possibilities could be studied within the doll play protocol. It would be interesting to see if the precursors to boredom are a sequence of redundant acts by the subjects. On the other hand, signs of disinhibition may be succeeded by behaviors we assume to be generally prohibited or have been specifically prohibited for the subject.

In summary, we are saying that there exist in the usual doll play data, possibilities for more elaborate and potentially more profitable analyses than have so far been made.

2. The above approach to the wish fulfillment and replication problem focuses on response measures. It is our belief that the study or manipulation of antecedent conditions also may be a fruitful tack.

Our first suggestion is to make use of detailed naturalistic information. A log of the child's experience for a day or two prior to doll play might be kept and the fantasy protocol compared with what we know occurred in the child's life. A very detailed log is represented by *One Boy's Day* (Barker & Wright, 1951). Such an approach is clearly inductive and simply provides a mass of data which may be scrutinized for simple correspondences or for more complex transformations between real life and doll play fantasy. For example, one could look at the ways in which an objectively described situation is filtered through the child's perceptions, and the results might provide clues to types of experiences which form the raw materials for wish fulfillment compared

to those types of experiences which are replicated with a high degree of fidelity.

The above naturalistic approach may point up variables, which, through experimental variation, will provide more substantial causal relationships between experiences and fantasy. For example, will a series of successes followed by a failure be fantasied as a success or failure? Does strongly goal oriented action that is not permitted consummation appear in doll play as goal achieved? Can a child be given a set to portray either wish fulfilling or replicative events?

In essence, we are suggesting that an experimental approach to the antecedents of children's fantasies has been tried very little and may provide substantial payoff. If significant antecedent manipulations are found, and their effects are potent and consistent across subjects, a more convenient response index to the two types of fantasy may appear. A case in point is the empirically derived scoring scheme for n Ach, which includes those categories of fantasy that respond consistently to experimental manipulation of arousal compared to neutral instructions.

Nature of Instigation in Doll Play

One of the presumed virtues of doll play is that the amorphousness of the stimulus situation would permit wide expression of "person" variables. Consequently, the preoccupations of the subject would be the major determinants of his fantasy responses. Recently, the contribution of the instigating stimulus itself has received serious attention in projection theory. For example, in the TAT measurement of need achievement the pictures were found to vary in the degree to which they elicited achievement imagery.

In doll play, we get the impression that the situation may be too broadly instigating for the purposes for which the technique is often used. Since this projective method is used to measure a wide variety of child behaviors, it is questionable if it is an equally appropriate measuring device for all of the variables. The data imply quite certainly that doll play is a useful device for measuring fantasy aggression. Beyond that, the incidence of other actions which may be coordinated to such motivational systems as dependency and achievement appear to be meagre. In other words, the home as the miniature situation is associated with so many kinds of behavior that the researcher cannot be sure that the actions in which he is interested will appear with sufficient frequency to be useful.

We can suggest two devices to narrow the spectrum of instigation. The first is to arrange a doll play setting which calls forth the specific behaviors upon which the study focuses. For example, several studies have used a school room rather than a house when the researcher was interested in school related behavior (Bach, 1945; Melville, 1959) and one study used a play yard setting (Bremer, 1947) to study play related behavior.

The second procedure focuses doll play even more narrowly, and may be thought of as analogous to the story completion technique. The experimenter presents the child with a problem and then permits the child to complete the action when the dolls are given to him. Lynn's recent structured doll play test follows this procedure; and the studies of prejudice in which the child is asked to make a choice of a doll is a second example of the focused method.

Summary

This paper surveyed the development and uses of doll play as a research tool. Besides methodological studies the findings in five areas of investigation which have used doll

play were summarized: aggression, stereotypy, doll preference, effect of separation from parents, and prejudice. Although certain groups of studies yield interrelated results, the use of this research tool has been so varied that the overall impression is of many disparate findings, in spite of the basic similarity in method. It is suggested that a conceptual difficulty underlying the studies has been the lack of distinction between wish fulfilling and replicative fantasies in children.

REFERENCES

ACKERMAN, N. W. Constructive and destructive tendencies in children. *Amer. J. Orthopsychiat.*, 1937, **7**, 301–319.

AMES, LOUISE B., LEARNED, JANET, MÉTRAUX, RUTH W., & WALKER, R. N. *Child Rorschach responses: Developmental trends from two to ten years.* New York: Hoeber, 1952.

AMMONS, CAROL H., & AMMONS, R. B. Research and clinical applications of the doll-play interview. *J. Pers.*, 1952, **21**, 85–90.

AMMONS, CAROL H., & AMMONS R. B. Aggression in doll-play: Interviews of two- to six-year-old white males. *J. genet. Psychol.*, 1953, **82**, 205–213.

AMMONS, R. B. Reactions in a projective doll-play interview of white males two to six years of age to differences in skin color and facial features. *J. genet. Psychol.*, 1950, **76**, 323–341.

AMMONS, R. B., & AMMONS, H. S. Parent preferences in young children's doll-play interviews. *J. abnorm. soc. Psychol.*, 1949, **44**, 490–505.

BACH, G. R. Young children's play fantasies. *Psychol. Monogr.*, 1945, **59**(2, Whole No. 272).

BACH, G. R., Father-fantasies and father-typing in father-separated children. *Child Develpm.*, 1946, **17**, 63–80.

BACH, G. R., & BREMER, GLORIA. Projective father fantasies of preadolescent, delinquent children. *J. Psychol.*, 1947, **24**, 3–17.

BARKER, R. G., & WRIGHT, H. F. *One boy's day.* New York: Harper, 1951.

BELLAK, L., & BELLAK, SONYA S. *Children's Apperception Test, manual.* (Rev. ed.) New York: CPS Co., 1950.

BIXLER, R. Nondirective play therapy. Unpublished master's thesis, Ohio State University, 1942.

BREMER, GLORIA. The effect of two fantasy environments on children's doll-play responses. Unpublished master's thesis, State University of Iowa, 1947.

BRYAN, CAROL. An experimental study of the dramatization of family life situations by young children. Unpublished master's thesis, University of Minnesota, 1940.

CARON, A. J., & GEWIRTZ, J. L. An investigation of the effects of the sex category of the interacting adult, chronological age (6, 8 and 10), and sex of child on aggressive (hostile) behavior in doll play. *Amer. Psychologist*, 1951, **6**, 307. (Abtract)

CLARK, K. B., & CLARK, MAMIE P. Racial identification and preference in Negro children. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947. Pp. 169–178.

CONN, J. H. A psychiatric study of car sickness in children. *Amer. J. Orthopsychiat.*, 1938, **8**, 130–141.

CONN, J. H. Children's reactions to the discovery of genital differences. *Amer. J. Orthopsychiat.*, 1940, **10**, 747–754.

DESPERT, J. L. A method for the study of personality reactions in preschool age children by means of analysis of their play. *J. Psychol.*, 1940, **9**, 17–29.

EMMERICH, W. A study of parental identification in young children. *Genet. psychol. Monogr.*, 1959, **60**, 257–308.

ERON, L. D. The use of theory in developing a design. Paper read at American Psychological Association, Washington, D. C., August 1958.

FINCH, HELEN M. Young children's concepts of parental roles. Unpublished doctoral dissertation, Florida State University, 1954.

FINCH, HELEN M. Young children's concepts of parent roles. *J. home Econ.*, 1955, **47**, 99–103.

FREUD, ANNA. Introduction to the technic of child analysis. *Nerv. ment. Dis. Monogr. Ser.*, 1928.

GEWIRTZ, J. L. An investigation of aggressive behavior in the doll play of young Sac and Fox Indian children and a comparison to the aggression of midwestern white preschool children. *Amer. Psychologist*, 1950, **5**, 294–295. (Abstract)

GOODMAN, MARY E. The genesis of race awareness and attitude. Unpublished doctoral dissertation, Radcliffe College, 1946.

GOODMAN, MARY E. *Race awareness in young children.* Cambridge, Mass.: Addison-Wesley, 1952.

GRAHAM, T. F. Doll-play phantasies of Negro and white primary-school children. Un-

published doctoral dissertation, University of Ottawa, 1952.

GRAHAM, T. F. Doll-play phantasies of Negro and white primary-school children. *J. clin. Psychol.*, 1955, **11**, 29–33.

HALNAN, HELEN H. A study of father-child relationships using a doll-play technique. Unpublished master's thesis, Stanford University, 1950.

HARDING, J., KUTNER, B., PROCHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 1021–1061.

HARTLEY, E. L., & SCHWARTZ, S. A pictorial doll-play approach for the study of children's intergroup attitudes. *Int. J. Opin. attitude Res.*, 1951, **2**, 261–270.

HEINICKE, C. M. Some effects of separating two-year-old children from their parents: A comparative study. *Hum. Relat.*, 1956, **9**, 105–176.

HENRY, J., & HENRY, ZUNIA. Doll play of Pilaga Indian children: An experimental and field analysis of the behavior of the Pilaga Indian children. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1944, No. 4.

HOLLENBERG, ELEANOR. A study of some effects of child-training practices on doll play behavior. Unpublished master's thesis, State University of Iowa, 1949.

HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.

HOLWAY, AMY R. Early self-regulation of infants and later behavior in play interviews. *Amer. J. Orthopsychiat.*, 1949, **19**, 612–623.

HONZIK, MARJORIE P. Sex differences in the occurrence of materials in the play constructions of preadolescents. *Child Develpm.*, 1951, **22**, 15–35.

ISCH, MARIA J. Fantasied mother-child interaction in doll play. *J. genet. Psychol.*, 1952, **81**, 233–258.

JEFFRE, MARIA F. D. Fantasied mother-child interaction in doll play. Unpublished doctoral dissertation, State University of Iowa, 1946.

JOHNSON, ELIZABETH Z. Attitudes of children toward authority as projected in their doll play at two age levels. Unpublished doctoral dissertation, Harvard University, 1951.

JOHNSON, L. C. The effect of father absence during infancy on later father-child relationship using a doll-play technique. Unpublished master's thesis, Stanford University, 1952.

KORNER, A. F. *Some aspects of hostility in young children*. New York: Grune & Stratton, 1949.

KRALL, VITA. Personality factors in accident-prone and accident-free children. Unpublished doctoral dissertation, University of Rochester, 1951.

KRALL, VITA. Personality characteristics of accident-repeating children. *J. abnorm. soc. Psychol.*, 1953, **48**, 99–107.

LEVIN, H. Effects of parental punishment on young children's overt and fantasy aggression. *Amer. Psychologist*, 1953, **8**, 387. (Abstract)

LEVIN, H. The influence of classroom control on kindergarten children's fantasy aggression. *Elem. sch. J.*, 1955, **55**, 462–466.

LEVIN, H. Permissive child rearing and adult role behavior. In D. E. Dulany, Jr., R. L. DeValois, D. C. Beardslee, & Marian R. Winterbottom (Eds.), *Contributions to modern psychology*. New York: Oxford Univer. Press, 1958.

LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll-play aggression. *Child Develpm.*, 1956, **27**, 135–153.

LEVIN, H., & TURGEON, VALERIE F. The influence of the mother's presence on children's doll-play aggression. *J. abnorm. soc. Psychol.*, 1957, **55**, 304–308.

LEVY, D. M. Use of play technique as experimental procedure. *Amer. J. Orthopsychiat.*, 1933, **3**, 266–277.

LEVY, D. M. Hostility patterns in sibling rivalry experiments. *Amer. J. Orthopsychiat.*, 1936, **6**, 183–257.

LEVY, D. M. "Control-situation" studies of children's responses to the difference in genitalia. *Amer. J. Orthopsychiat.*, 1940, **10**, 755–762.

LYNN, D. B. Development and validation of a Structured Doll Play Test for children. *Quart. Bull. Ind. U. Med. Cent.*, 1955.

LYNN, D. B., & LYNN, ROSALIE. The Structured Doll Play Test as a projective technique for use with children. *J. proj. Tech.*, 1959, **23**, 335–344.

LYNN, D. B., & SAWREY, W. L. The effects of father-absence on Norwegian boys and girls. *J. abnorm. soc. Psychol.*, 1959, **59**, 258–262.

LYNN, ROSALIE. A study of the responses of four- and six-year-olds to a structured doll-play test. Unpublished master's thesis, Purdue University, 1955.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MEISTER, D. Adjustment of children as reflected in play performance. *J. genet. Psychol.*, 1948, **73**, 141–155.

MELVILLE, CLARA P. A study of overt and fantasy expressions of variables related to young children's motivation toward working industriously in school. Unpublished

doctoral dissertation, Stanford University, 1959.

MILLER, H., & BARUCH, DOROTHY W. A study of hostility in allergic children. *Amer. J. Orthopsychiat.*, 1950, **20**, 506–519.

MUSSEN P., & DISTLER, L. Masculinity, identification, and father-son relationships. *J. abnorm. soc. Psychol.*, 1959, **59**, 350–356.

PHILLIPS, RUTH. Doll play as a function of the realism of the materials and the length of the experimental session. *Child Develpm.*, 1945, **16**, 123–143.

PINTLER, MARGARET H. Doll play as a function of experimenter-child interaction and initial organization of materials. *Child Develpm.*, 1945, **16**, 145–166.

PINTLER, MARGARET H., PHILLIPS, RUTH, & SEARS, R. R. Sex differences in the projective doll play of preschool children. *J. Psychol.*, 1946, **21**, 73–80.

RABBAN, M. Sex-role identification in young children in two diverse social groups. *Genet. psychol. Monogr.*, 1950, **42**, 81–158.

RADKE, MARIAN J. The relation of parental authority to children's behavior and attitudes. *Inst. Child Welf. monogr. Ser.*, 1946, **22**.

RADKE, MARIAN J., & TRAGER, HELEN G. Children's perceptions of the social roles of Negroes and whites. *J. Psychol.*, 1950, **29**, 3–33.

ROBINSON, ELIZABETH F. Doll play as a function of the doll family constellation. *Child Develpm.*, 1946, **17**, 99–119.

ROSENZWEIG, S., & SHAKOW, D. Play technique in schizophrenia and other psychoses. *Amer. J. Orthopsychiat.*, 1937, **7**, 36–47.

RYDER, JOYCE M. Aggression with balloons, blocking, and doll play. In Lois M. Stolz and collaborators, *Father relations of war-born children*. Stanford: Stanford Univer. Press, 1954. Pp. 212–243.

SCOTT, R. G. Projective parental fantasies of parent-separated children and children in families. Unpublished EdD dissertation, Teachers College, Columbia University, 1954.

SEARS, PAULINE S. Doll-play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**(6, Whole No. 323).

SEARS, PAULINE S., & PINTLER, MARGARET H. Sex differences in doll-play aggression. *Amer. Psychologist*, 1947, **2**, 420. (Abstract)

SEARS, PAULINE S., & STAFF OF THE LABORATORY OF HUMAN DEVELOPMENT, HARVARD UNIVERSITY. Child-rearing factors relating to playing of sex-type roles. *Amer. Psychologist*, 1953, **8**, 431. (Abstract)

SEARS, R. R. Influence of methodological factors on doll-play performance. *Child Develpm.*, 1947, **18**, 190–197.

SEARS, R. R. Relation of fantasy aggression to interpersonal aggression. *Child Develpm.*, 1950, **21**, 5–6.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.

SEARS, R. R., PINTLER, MARGARET H., & SEARS, PAULINE S. Effect of father separation on preschool children's doll-play aggression. *Child Develpm.*, 1946, **17**, 219–243.

SIMPKINS, VERA A. A comparison of verbal and motor fantasies in children. Unpublished master's thesis, Clark University, 1948.

SPERRY, MARGARET. An experimental study of the effects of disapproval of aggression on doll play. Unpublished master's thesis, State University of Iowa, 1949.

STAMP, I. M. An evaluation of the Driscoll Playkit used with incomplete stories as an instrument for the diagnosis of personality. Unpublished EdD dissertation, Teachers College, Columbia University, 1954.

STEVENSON, H. W., & STEWART, E. C. A developmental study of racial awareness in young children. *Child Develpm.*, 1958, **29**, 399–409.

TUDDENHAM, R. D. Studies in reputation: I. Sex and grade differences in school children's evaluation of their peers. II. The diagnosis of social adjustment. *Psychol. Monogr.*, 1952, **66**(1, Whole No. 333).

WALSH, ANN M. *Self-concepts of bright boys with learning difficulties*. New York: Teachers Coll., Columbia Univer., Bureau of Publications. 1956.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

WINSTEL, BEULAH. The use of a controlled play situation in determining certain effects of maternal attitudes on children. *Child Develpm.*, 1951, **22**, 299–311.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

WURTZ, K. R. The expression of guilt in fantasy and reality. Unpublished doctoral dissertation, Stanford University, 1957.

WURTZ, K. R. Some theory and data concerning the attentuation of aggression. *J. abnorm. soc. Psychol.*, 1960, **60**, 134–136.

YARROW, L. J. The effect of antecedent frustration of projective play. *Psychol. Monogr.*, 1948, **62**(6, Whole No. 293).

(Received October 31, 1960)

Psychological Bulletin
1962, Vol. 59, No. 1, 57–69

A COMPARATIVE-DEVELOPMENTAL APPROACH TO SCHIZOPHRENIA

ALFRED E. GOLDMAN[1]
Norristown State Hospital, Pennsylvania

There is a growing number of social and biological scientists who feel the need for a comprehensive theory of behavior—a theory of which schizophrenia in particular, or psychopathology in general, is only one facet. The theory should be broad enough to encompass data from such apparently diverse fields as anthropology, phylogenesis, human development, and states of lowered consciousness. Data from all of these areas contribute to our understanding of human behavior, and it would seem that the law of parsimony would be better served if these data could be subsumed under the same concepts and interpreted in terms of a common set of principles.

This paper attempts to outline a comparative-developmental approach to schizophrenia. It is comparative in that it relates data from the study of schizophrenia to many different fields of inquiry. It is developmental insofar as it is suggested by, and draws its basic facts from developmental studies—the development from conception to birth, the development from childhood to adulthood, the development from the single-celled organisms to man, and from developmental studies of human cultures.

For the particular organization of the approach to schizophrenia presented here, the author accepts responsibility; the original formulation of the comprehensive comparative-developmental theory is that by Heinz Werner (1940) and his coworkers at Clark University.

Werner's comparative-developmental approach aims at viewing the total behavior of all organisms in terms of a common set of developmental principles. It is his belief that such an approach is fruitful in coordinating, within a single descriptive framework, psychological phenomena observed in phylogenesis, ontogenesis, ethnopsychology, and psychopathology. This paper confines itself to what this theoretical position has had to contribute to an understanding of schizophrenia. It attempts to indicate the comprehensiveness and heuristic value of the approach without, however, attempting to present an exhaustive review of the large body of relevant research.

Behavior proceeds through given stages in its development. A formal similarity obtains between the organization and structure of processes in young children, in organisms low on the phylogenetic scale, in human adults of technologically backward societies, and in certain states of lowered consciousness in educated normal adults of technologically advanced societies. In order for developmental theory to encompass schizophrenic processes it requires the introduction of constructs which suggest a parallelism of various aspects of schizophrenia with developmental patterns in all of these spheres of inquiry, but especially with development in childhood. To this end developmental theorists have introduced the concept of "regression." The progression seen in the normal

[1] Presently with National Analysts, Incorporated, 1015 Chestnut Street, Philadelphia, Pennsylvania.

course of development is reversed in pathology; thus, in schizophrenia we may expect to find a regression in the direction of greater primitivization of process.

A frequently raised objection to developmental theory is that it seeks only generic similarities between various groups and tends to ignore their differences.

Exploration of developmental theory does require seeking for systematic patterns of generic similarities in cognitive performance among certain groups. Thus focused on similarities, developmental theorists have not always taken explicit account of specific differences that have appeared between groups.

The heuristic value of such an approach has already been demonstrated by the considerable number of investigations that have been provoked by or conducted under the purview of development theory. Its clinical value is suggested by its contributions to psychodiagnostic testing, in particular to the scoring and interpretation of the Rorschach technique. Genetic theory does not question that differences exist between the child and adult schizophrenic. It does hold that similarities in cognitive structure exist between young children and adult schizophrenics both of which are exemplifications of an ideal construct, namely, developmental primitivity.

A word now about the use of the term "primitive" (Werner & Kaplan, 1956). Much of the criticism leveled at the use of this term is based on the assertion that it is moralistic in character and thus has little place in scientific endeavor. No such evaluative connotation is intended. While "primitivity" is not evaluative in this moralistic sense, it is evaluative in that it may either impede or facilitate attainment of certain goals or states. Primitivity pertains to the psychologically prior stages of development. In essence the concept of primitivity is a theoretical construct referring to a kind of cognition characterized by developmentally early processes. Processes that appear early in the development sequence—that is, early in childhood, or early in the temporal development of an idea—are more primitive than those which appear later in the sequence.

The term "regression" as used by Werner (1940) refers to the structural re-emergence of developmentally lower levels of functioning as the more advanced and more recently developed levels are disorganized. Regression in this sense differs in emphasis from the meaning given this term by psychoanalytic orthodoxy[2] which focuses on impulses and the methods by which these are gratified and controlled. While psychoanalysis has emphasized the *function* and *content* of psychopathology, the developmental approach considers only the *formal structure* of psychopathological processes.

By similarity in *process* between childhood and pathological primitivization reference is made to structural similarity, not to similarity in content. The regressed adult is, of course not a child; rather, similar organizations or forms of process are identifiable in both. Our interest here is not primarily in *what* children or schizophrenics think or perceive, but rather, *how* they think or perceive. Schizophrenia thus is seen as a regression in cognitive processes; that is, it is conceived as a reversal of those patterns of thinking, perceiving, and so on, which are encountered in the normal course of development. Further, developmental theorists are not con-

[2] Although Freud considered ego regression as well as impulse regression, many psychoanalytic practitioners are inclined to overemphasize the latter at the expense of the former.

cerned with the nature of the conditions that have caused the regressed behavior or the historical antecedents of such conditions. Rather they focus on the structural or formal consequences of these predisposing experiences.

It should be made clear that the psychoanalytic and the comparative-developmental approaches are not mutually exclusive; rather, they focus on different aspects of schizophrenia (Arieti, 1955). Each may be clinically useful and theoretically productive. Devoting attention in this paper to the structural point of view does not attribute less value or validity to the psychodynamic viewpoint. Where the psychodynamic approach is particularly helpful in therapy, the structural approach is useful in developing hypotheses, describing developmental phenomena within a consistent framework, and—most important to the clinician—it provides a gauge by which psychopathological states and modifications in those states may be assessed and understood in terms of developmental criteria (Siegel, 1953). The concept of schizophrenia which is proposed here proceeds from a basic developmental principle; wherever development takes place it initiates in a globality or lack of differentiation and becomes increasingly more differentiated, terminating in a state of integration. The development of motor coordination may serve to illustrate this developmental principle.

When stimulated, the newborn typically reacts with mass nondirected motor activity. In the normal course of maturation, this mass action becomes more focalized and better directed with respect to the stimulating agent. That is, from the total involvement of the whole body emerges a differentiated activity of certain parts of the body—arms, legs, head, etc. These now differentiated movements become integrated into a single smooth-flowing response in which all parts of the body may participate appropriately in achieving a goal or solving a task.

Now let us turn to the separate functions that this approach encompasses.[3] In each case the comparison will be made between human ontogenesis and schizophrenia.

Emotional Behavior

Ontogenetic changes in emotional behavior proceed along, at least, three continua: (*a*) From overt motor expression of emotion to increasingly more internalized experience of emotion. Crying (Bayley, 1932), and other motor activity decreases with age. (*b*) From globality of emotional experience to greater differentiation (Bridges, 1932). At first there are only undifferentiated affective states of relative excitement or quiescence. With development there is greater specificity of emotion. For example, global negative affect becomes more differentiated into increasingly more subtle nuances, such as hate, despise, contempt, dislike, etc. (*c*) From lability of emotional experience to increased stability. In the young child there is characteristically momentary change in the nature of his emotional experiences and its expression (Jersild, 1939). What starts out as a laugh may end up in bitter tears or vice versa. Crying can be quickly changed to giggling by a well intentioned and well placed tickle.

In accordance with the regression hypothesis, in schizophrenia there is the expectation of a reversal in each of these three progressions:

1. In the acute stage of the illness, before chronicity becomes manifest

[3] A comprehensive survey of developmentally oriented research in childhood may be found in Werner (1946).

in affective blunting, emotion is uncontrolled; impulse is expressed overtly without adequate intellectual intervention. Not only is the expression of affect likely to be more public, but there is an increase in the degree of motor involvement. Thus, the motoric hyperactivity of the excited schizophrenic and the motoric hypoactivity of the chronic "burnt-out" schizophrenic both exhibit the degree to which the emotional state is syncretically (Werner, 1940) fused in its expression with the motoric system. Although the affective and motoric are never wholly independent (Wolff, 1943) of each other, the immediacy, directness, and overtness of this relationship tends to increase in schizophrenia.

2. The increasing differentiation and subtlety of feelings seen in ontogenesis is reversed in schizophrenia. Clinical practice, in particular experience with the projective techniques, reflects the dedifferentiation of feelings. Aggressive and sexual components are not infrequently fused into an indistinguishable whole. Even more striking is the blatant admixture of positive and negative impulses.

3. Though perhaps not to the same degree, the emotional experience of the acute schizophrenic is similar to that of the young child's in that it, too, is highly labile and unpredictable.

Perception

The progression from globality to differentiation to integration is perhaps best seen in perception. For the neonate and very young child the visual field is not well organized or structured. Figure and ground, contours, patterns of light and shadow, movement, all merge into an undifferentiated perceptual mass, or in William James' classic terminology, "a blooming, buzzing confusion." From this globality emerges stages of increasingly differentiated perception. Here visual patterns acquire object-properties, with definitive contours and localized in three-dimensional space. This development then terminates in a stage in which these differentiated aspects of the perceptual field are integrated, or synthesized, into a single meaningful percept (Werner, 1940).

This developmental sequence has been corroborated by a number of experiments, the most convincing of which have used the Rorschach blots as stimulus material (Hemmendinger, 1953). Use of this technique reveals the following changes to take place with increasing age.

Three-year-olds are whole-perceivers; they see few details and their perception is best described qualitatively in terms of their undifferentiated character. Four- and 5-year-olds react less in terms of wholes and more often notice and comment on the parts. At 6 years another, and distinct, change occurs: an abrupt and marked increase in perceptual responses to the small and rarely noticed areas in the blots. This attraction to tiny details is interpreted as an intensification of the development of differentiation. At 9 years begins the final phase of perceptual development—that of synthesis and integration. This final phase terminates in the appearance of predominantly synthesizing activity. In the integrated whole response, the blot is perceptually articulated and then reintegrated into a well differentiated unified whole.

Having considered perceptual development in children we would expect, according to the regression hypothesis, a reversal of this pattern in schizophrenia. Further, we would expect that the greater the pathology the more immature the perception.

Experiments, particularly those by Friedman (1953) and Siegel (1953),

reveal the following relationships in perceptual function between schizophrenics and children:

With respect to the developmentally immature response, there exists no significant difference between children and schizophrenics, and both groups differ significantly from normal adults. The same is true of the most advanced percepts. The integrated whole response discriminates each of the three groups from each other. Thus, these findings justify the conclusion that schizophrenics, in some respects, respond perceptually in a manner similar to that of children, and in other aspects, they occupy an intermediate position between normal adults and children. This may be understood in terms of the hypothetical construct of regression. In this regard regression seems evident, but it is not of such a total nature as to completely eradicate the history of the individual who has once operated on a higher developmental level.

Now, what may be said regarding the schizophrenic subtypes? There is little or no evidence on which to discriminate the perceptual functioning of the hebrephrenics and catatonics from each other, and no work has been done with simple schizophrenics. However, developmentally comparing paranoid schizophrenics with the combined hebrephrenic and catatonic group (Siegel, 1953) we find the following: while the perception of paranoid schizophrenics is typically fractionated and fragmented with emphasis on perceptual analysis, resembling the performance of children from 6 to 10, that of the hebrephrenic and catatonic schizophrenics is characteristic of the global, amorphous perceptual activity of 3–5 year old children.

Comparative-developmental theory thus permits the location of catatonics, hebrephrenics, and paranoids on a developmental scale. In all aspects of cognitive functioning, in addition to perception, paranoid schizophrenics are expected to perform more like the normal adult than the catatonic or hebrephrenic schizophrenic. It does not, however, attempt to state the conditions which facilitate or inhibit the depth of regression in these diagnostic categories. At this stage in its development the theory has paid relatively little attention to motivational aspects of schizophrenia. Among clinical practitioners this conceptual vacuum has been filled by psychodynamic theories.

There are other aspects of perceptual development and regression that are instructive here:

The extreme lability that we see in primitive emotional behavior is also seen in the perceptual sphere. Those who have worked intensively with schizophrenics or with young children cannot avoid being impressed by the extreme lability of their attention. This, in both the child and in the schizophrenic, may be attributable to a kind of perceptual passivity in which competing stimuli have equal potential for evoking a perceptual response. This notion of stimuli equipotentiality may be useful in understanding the severe stimulus boundedness of the child and schizophrenic.

The child is stimulus bound in that the stimulus *must* be attended to. An infant's eyes *must* follow the hand that goes before it. His hand must grasp the object that is placed in it.

The schizophrenic is similarly stimulus bound. Stimuli that compete for a perceptual response cannot be adequately discriminated in terms of their relevance to a task. Thus, the schizophrenic complains of a rapidly shifting, kaleidoscopic world. A patient seen by the author complained continually that he could not attend

to anything for very long because everything and anything disrupted his thoughts. Apparently irrelevant details demanded his attention: a noise outside, lights passing by at night, an apparently random thought, or a bodily sensation had equal demand on his attention as the topic being discussed or the task at hand. This extreme interpenetration of the schizophrenic's attention and thought by apparently random stimuli is a well known phenomenon and has been well described by Cameron (1939), Kasanin (1944), and others.

Learning

The developmental approach to learning derives from the notion that development is characterized by qualitatively different processes and modes of organization, rather than by simply quantitative variations in process. This approach is therefore in opposition to those theoretical orientations which view learning as reduceable to a single process. Developmental theory does not conceive of any one process as being paradigmatic of the whole range of human learning. A view which reduces all learning to a single process conceives of the adult as having available *more* response alternatives than the child. A genetic point of view conceives of the adult and child as utilizing *different* processes which may not be distinguishable in terms of efficiency or achievement.

Developmental theorists thus seek to understand the nature of human learning through the exploration of qualitatively distinct organizational stages. Such an exploration was undertaken in a recent study by Goldman and Denny (in press). They presented two kinds of learning tasks to children 5–14 years old. Performance in the first learning task depended on apprehending the regular pattern of the pre-established program (response to two switches in a right-right-left-left sequence). Performance in this task increased steadily with age and IQ. In the second task rewards were received according to a predetermined, random "probability" program in which one response was rewarded 25% of the time and the other response was rewarded 75%. Performance in this task was essentially invariant with age and IQ with the trend somewhat favoring the younger children. Insofar as these developmental curves were strikingly different they were interpreted as indicating that the performances on the two learning tasks reflected different processes. Insofar as the sequential, or "recursive," task required an active seeking for a general rule for its solution, it was interpreted as requiring a more advanced mode of functioning than that on the probability or "stochastic" task which permitted a more passive orientation to the task in that it did not provide for such an easily generalizable solution.

A third learning process that may represent the most primitive level for humans is classical conditioning, in which the stimulus is presented wholly at the discretion of the experimenter and the response is usually of a physiological or reflexive nature. Developmental studies of classical conditioning suggest that conditioned responses can be established very early in life and indeed that young children can be more easily conditioned than older children and adults (Jones, 1928, 1930a, 1930b; Kasatkin & Levikova, 1935; Mateer, 1918; Razran, 1933, 1935). The developmental primitivity of classical conditioning is further suggested by studies which indicate that susceptibility to conditioning is enhanced in states of lowered conscious-

ness (Leuba, 1940, 1941; Scott, 1930).

Thus, at least three modes of learning are suggested which, in the order from most primitive to most advanced, are: learning by classical conditioning, stochastic learning (instrumental conditioning), and recursive learning (problem solving). The first level appears to be characteristic of the learning of very young children and of infrahuman animals. Here the learner is a kind of passive "victim" of his environment in that he does little of an active nature to learn; learning, the pairing of stimuli and response, is imposed upon him.[4] The second learning mode is distinguished from the first in that the learner is active or "instrumental" in the learning process, yet the learning process is essentially by rote. In this learning mode young children and adults do equally well, as do subjects of varying intelligence. The third learning mode is not only the most active in that there is a deliberate seeking for order and regularity, but there is a vigorous development and testing of solution hypotheses. This learning mode favors older and more intelligent subjects.

With growth—phylogenetic and ontogenetic—classical conditioning is less adaptive and recedes to the background until called upon when the task situation calls for no more profound level of intellection. The other modes of learning emerge later to better serve the individual's needs.

In schizophrenia it is proposed that this development is reversed, with sequential learning and other forms of complex learning situations being effected most and classical conditioning ascending in relative importance.

Schizophrenics have been found to be more readily conditioned than normals in relatively simple situations in which the response alternatives are limited and the response reflexive. This has been demonstrated for the knee jerk (Pfaffman & Schlossberg, 1936), the psychogalvanic response (Mays, 1934; Shipley, 1934), and eyeblink (Spence & Taylor, 1953). Schizophrenics have also been shown to exceed neurotics in eyeblink conditioning (Taylor & Spence, 1954). However, since some studies have failed to demonstrate the greater conditionability of schizophrenics over normals (Howe, 1958; Paintal, 1951), the question is raised as to what stimulus conditions enhance the establishment of the conditioned response in schizophrenics as compared to normals.

In accordance with the regression hypothesis, the increase in susceptibility to conditioning in schizophrenia should be accompanied by a decrement in performance of complex tasks. By "complex" task is meant tasks which permit wide response alternatives, among which are many irrelevant ones, and in which an active role of the learner is required. Schizophrenics have been found to perform poorly relative to the performance of control normals in these complex tasks (Cameron, 1939; Hanfmann, 1939; Hanfmann & Kasanin, 1942; Rapaport, 1945).

The increased conditioning performance and the decreased performance in complex tasks, in schizophrenia as compared to normals, has been interpreted by Mednick (1958) and other learning oriented theorists (e.g., Taylor & Spence, 1954) in terms of the effect of drive intensification (anxiety) on the response strength of the conditioned response. A difficulty with this type of Hullian interpretation is that it fails to take into account developmental data. The superior performance of children

[4] A similar viewpoint was expressed by Gesell (1938).

and infrahuman animals relative to normal adults in conditioning experiments can hardly be incorporated within such a theoretical framework unless one postulates the existence of a heightened drive state in these more primitive organisms. Genetic theory offers the parsimonious incorporation of data from all of these areas within a single theoretical structure.

When a stable stimulus-response relationship has been established the response may be elicited by other stimuli similar in some manner to the initial stimulus. This is stimulus generalization.

The genetic principle that differentiation proceeds from an initial stage of globality would suggest that in development stimulus generalization would decrease. Reiss (1946) found that young children tend to generalize readily to homophones but this tendency disappears at about 11 years of age. Mednick and Lehtinen (1957) found that amount of stimulus generalization reactivity, measured along a visual-spatial dimension of similarity, was significantly greater for younger children (7–9 years) than for older children (10–12 years).

The expectation then would be that in schizophrenia stimulus generalization would be higher than in normals of comparable intelligence. A number of studies testify that this is so (Cameron, 1938; Garmezy, 1952; Mednick, 1955).

Thinking and Language

Thinking and language may be investigated from the vantage of many dimensions. Three which appear to the author to be most central and inclusive are the development from idiosyncrasy to consensuality of concepts, from lability to stability of concepts, and from contextualization to autonomy of concepts.

The development from idiosyncrasy to consensuality refers to the increasingly more public and predictable thinking of which the child becomes capable as he grows older (Pollack, 1953; Werner & Kaplan, 1952). Thus, the agreement in the meaning of words among members of a given speech community increases with age. Children, in contrast to adults, use words in a private, highly individualistic manner (Hayakawa, 1954).

In psychopathological regression the development toward greater consensuality in thinking is reversed. Idiosyncratic thought then reduces the schizophrenic to virtual social isolation (Cameron, 1938; Goldman, 1960).

The second dimension is the development from lability to stability of concepts. In the young child concepts are typically labile (Pollack, 1953). The nature of the concept changes rapidly and in a seemingly capricious manner (Eng, 1931).

An example from performance on the Object Sorting Test (Rapaport, 1945) may serve to illustrate concept lability. The test consists of a number of everyday, common objects that are placed on a desk before the subject. The typical adult, when asked to place these objects into meaningful groups so that the objects within any one group belong together, will form objects into groups according to their color, or material, or perhaps their use. A subject may pick out all red objects and put them together, or all wooden objects, or all tools. Young children will frequently switch the relationship in a very labile manner (Reichard, Schneider, & Rapaport, 1944). Thus, a young child will select first a red ball and then this is placed with a red plate, the two objects having redness in common. Then a toy knife is selected because it goes on the table,

too, like the red plate, and then pliers are chosen because it is metal like the knife, and then a pipe because "the workman uses the pliers and smokes a pipe."

Similar chain concepts are developed by schizophrenics in the same task situations. The response of a young schizophrenic girl in a task involving a linear schematization technique may serve as an illustration of the extreme equivocality, or lability, of the relationship between the symbol and the meaning it symbolizes (Goldman, 1960). Linear schematization requires the subject to represent a word, in this case a mood term, by drawing a line. The subject is asked to draw an "angry" line, or a line that expresses the word "misery," and so on. This subject was asked to draw a line that represented the word "healthy." She drew a series of different lines. When asked what there was in the lines she drew that suggested health she responded: "A seven upside down, lightning going up, the medusa, and this is the medical sign of health." While the patient could not clarify the way in which all of these concepts are related to health, the response invites speculation about the way each thought was related to the one that preceded it. While the experiment was in progress she was drinking 7-Up and remarked that it was "good for you." Lightning going up may represent a denial of the destructive (i.e., unhealthy) effects of lightning. The medusa may be related to "the medical sign of health" (the caduceus) by clang association, or by the snakes which are common to both.

In the extreme case concept lability may be reflected in one word or symbol subsuming not only different concepts but opposite ones. This has been established in dreams (Jones, 1913), in archaic language (Freud, 1950), and also in schizophrenia (Goldman, 1960).

The equivocal nature of symbol meaning in childhood and in schizophrenia appears to be determined by the close bond between the symbol and some particular situation, event, or person with which it is associated. This is the third dimension—the development from contextualization to autonomy of a concept. Concepts in childhood are determined by personally relevant experience (Binet, 1916; Chodorkoff, 1952; Feifel, 1949; Hayakawa, 1954; Kasanin, 1944; Terman, 1916). A newspaper, for example, may be defined as "what the paper boy brings and you wrap the garbage with it" (Hayakawa, 1954, p. 80). With growth these concepts become increasingly independent or autonomous of these personally meaningful contexts (Werner, 1940; Werner & Kaplan, 1950, 1952).

In schizophrenia we expect the reverse of this development: concepts should become increasingly less autonomous and more contextualized. There is extensive evidence—clinical and experimental (Arieti, 1948; Baker, 1953; Cameron, 1938; Goldman, 1960; Kasanin, 1944) that this is so. The vocabulary test performances lend further credence to the statement that in comparison to normals, schizophrenics tend to use words in terms of their concrete functions rather than in terms of abstract autonomous properties (Chodorkoff, 1952; Feifel, 1949; Harrington, 1954; Yacorzyncki, 1941).

This regression may be illustrated by referring again to linear schematization. A group of schizophrenics were asked to represent the meaning of a word in a line. Then inquiry was made into the relationship between the line and the word it expressed. Typically, the line was justified in

terms of some personally relevant experience. For example, the word "gentle" was represented by a patient as a hay stack when she replied to the inquiry with "lying in the hay is gentle." Another patient drew two lines which she said represented the path taken by the hand of a mother "gently" caressing a child. Still a third patient represented the word "gentle" with a leaf, which "is 'gently' blowing in the breeze." Gentleness in all of these cases is represented by unique personal experiences and associations. Similarly, in the Object Sorting Test, schizophrenics are more inclined than normals, to relate objects in a highly personal manner—"All of these things were in my mother's house" or "I think they are all pretty."

Thus, three dimensions of concepts are suggested. Underlying the first, idiosyncrasy-consensuality, is the increasing stability of concepts. A concept must be stable in reference before it can be public, or consensual. Underlying, in turn, the second dimension, is the contextuality-autonomy dimension. If a concept has meaning only in terms of personal contexts, its reference will be as labile as one's personal experiences, and therefore not available for use as a vehicle for social interaction.

The second and third dimensions both reflect the developmental progress from globality to differentiation, and its dedifferentiation in psychopathological regression. To the extent that a concept is labile, or in the extreme, in that it encompasses opposite meanings, it is undifferentiated. In schizophrenia the vehicles of thinking and communication become progressively dedifferentiated in that they, the symbol and referrent, are not related in a stable manner. With regard to contextualization it may be said that the more autonomous a meaning, the more it is differentiated from a particular context. Thus, in development there is progressive meaning-context differentiation, while in schizophrenia meaning and context are dedifferentiated.

Normal subjects more frequently reflect less situational meanings and attempt to represent some essential quality of gentleness. The word "gentle" is typically symbolized by normals by a light curved line, expressing the "soft," "light" aspects of "gentle." The autonomous meaning of a word is essential in that it abstracts from each of the many situations with which it is associated (lying in hay, mother caressing child, etc.), a commonity that each shares. The essential meaning of a concept is abstracted from but is relatively autonomous of concrete contexts.

Socialization

In the development of social behavior we again see the increasing differentiation out of the state of globality which terminates in integration. We have little reason to believe that in the neonate the self is distinguished from others. According to psychoanalytic theorists the mother, her breast, her voice, the warmth of her body, the sensations from within the infant's own body, are an indistinguishable whole. With development, there is an increasing awareness of the self as an entity.

The development in social integration is seen in patterns of play (Buehler, 1935; Loomis, 1931). At first, young children play in isolation with their hands, feet, or other objects. Later, children prefer to play in the presence of other children—not *with* other children, but in "parallel" play. Differentiation has taken place, with this first step toward integration and will eventually lead to

genuine interpersonal interaction.

This development toward social integration is also seen in the increasing complexity of the social groups, and in their increasing stability (Zaluzhni, 1930).

In schizophrenia we find similar processes, except in reverse. On the ward we can see interaction representing all of these phases. The suspicious, hostile paranoid that still seeks social interaction; the hallucinating, babbling chronic schizophrenic that somehow still prefers to hallucinate and babble in the presence of others, although not with or in concert with others; and finally, the totally regressed isolate who withdraws into the social vacuum of a corner of the ward and devotes himself to his own bodily sensations.

Motor Functions

One of the most striking developments to take place in the motor sphere is the increase in the implicitness of motor activity. Vicarious movements replace overt activity in reasoning, problem solving is less vocal and more silent, motion in general is less gross.

Relative to the massive debilitation in other spheres there is relatively little motor involvement in schizophrenia. It is only in the most severe regression that motor impairment is found, such as in catatonic *cerea flexibilitas*, and in the hyperactivity and restlessness that sometimes characterizes the acute stage of schizophrenia. In chronic schizophrenia, too, there is frequently evidence of incessant repetitive movements of head, trunk, or limbs.

The fact that there is little motor involvement in schizophrenia, except in severe cases, is consistent with Hughlings Jackson's principle that those functions which are the latest to develop are the first to be impaired in pathology. Since motor functions are amongst the first to develop in infancy, we would therefore expect impairment in this sphere to develop last.

There are other dimensions that have not been considered. In each of those that have been discussed focus has been on structural similarities between young children and schizophrenic functioning. Such similarities in process are also distinguishable in primitive cultures and in states of lowered consciousness, such as dreams, drug states, and hypnogogic conditions.

A comparative-genetic approach is fruitful in our effort to understand the essential nature of schizophrenia because it seeks to expose process rather than assess achievement and it is an approach in which structure is no less important than content and function.

Although a structural point of view has been central in the systems of some theorists for some time (Arieti, 1957; Munroe, 1955; Rapaport, 1951a, 1951b), psychoanalytic orthodoxy has not given sufficient attention to structural elements until recently. Having concerned itself in its early development predominantly with primary process, psychoanalysis is now turning increasingly more to a consideration of secondary process. Merton Gill (1959) has formalized this emphasis of the structural point of view in psychoanalysis.

This more energetic psychoanalytic consideration of ego functions, and the theoretical approach that has been offered in this paper have a similar goal—the formulation of a comprehensive theory of human behavior. Such genetic approaches remind us that in our consideration of the schizophrenic, oral deprivation is a no more significant datum than is the inability to conceive of square things in terms of their squareness.

REFERENCES

ARIETI, S. Special logic of schizophrenia and other types of autistic thought. *Psychiatry*, 1948, **11**, 325–338.

ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.

ARIETI, S. The two aspects of schizophrenia. *Psychiat. Quart.*, 1957, **31**, 403–416.

BAKER, R. W. The acquisition of verbal concepts in schizophrenia: A developmental approach to the study of disturbed language behavior. Unpublished doctoral dissertation, Clark University, 1953.

BAYLEY, NANCY. A study of the crying of infants during mental and physical tests. *J. genet. Psychol.*, 1932, **40**, 306–329.

BINET, A., & SIMON, T. *The development of intelligence in children*. Vineland, New Jersey: Training School, 1916.

BRIDGES, K. Emotional development in early childhood. *Child Develpm.*, 1932, **3**, 324–334.

BUEHLER, CHARLOTTE. *From birth to maturity*. London: Kegan Paul, 1935.

CAMERON, N. S. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**(1, Whole No. 221).

CAMERON, N. S. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, **85**, 1–24.

CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. *J. consult. Psychol.*, 1952, **16**, 43–48.

ENG, H. *The psychology of children's drawings*. London: Kegan Paul, 1931.

FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. psychol. Monogr.*, 1949, **39**, 151–204.

FREUD, S. The antithetical sense of primal words. In E. Jones (Ed.), *Collected papers*. Vol. 4. London: Hogarth, 1950.

FRIEDMAN, H. Perceptual regression in schizophrenia: An hypothesis suggested by the use of the Rorschach test. *J. proj. Tech.*, 1953, **17**, 171–185.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **20**, 253–267.

GESELL, A. The conditioned reflex and the psychiatry of infancy. *Amer. J. Orthopsychiat.*, 1938, **8**, 19–30.

GILL, M. The present state of psychoanalytic theory. *J. abnorm. soc. Psychol.*, 1959, **58**, 1–9.

GOLDMAN, A. Symbolic representation in schizophrenia. *J. Pers.*, 1960, **28**, 293–316.

GOLDMAN, A. Classification of sign phenomena. *Psychiatry*, in press.

GOLDMAN, A., & DENNY, J. Ontogenesis of choice behavior in stochastic and recursive programs. *J. genet. Psychol.*, in press.

GOLDSTEIN, K. *The organism*. New York: American Book, 1939.

HANFMANN, E. Thought disturbance in schizophrenia as revealed by performance in a picture completion test. *J. abnorm. soc. Psychol.*, 1939, **34**, 249–264.

HANFMANN, E., & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Mongr.*, 1942, No. 67.

HARRINGTON, R., & EHRMANN, J. Complexity of response as a factor in the vocabulary performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1954, **49**, 362.

HAYAKAWA, S. I. *Language in action*. New York: Harcourt, Brace, 1954.

HEMMENDINGER, L. Perceptual organization and development as reflected in the structure of the Rorschach test response. *J. proj. Tech.*, 1953, **17**, 162–170.

HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic subjects. *J. abnorm. soc. Psychol.*, 1958, **56**, 183–189.

JERSILD, A. T. *Child psychology*. New York: Prentice-Hall, 1939.

JONES, E. *Papers on psychoanalysis*. London: Bailliere, 1913.

JONES, H. E. Conditioned psychogalvanic responses in infants. *Psychol. Bull.*, 1928, **25**, 183–184.

JONES, H. E. The galvanic skin reflex in infancy. *Child Develpm.*, 1930, **1**, 106–110.

JONES, H. E. The retention of conditioned emotional responses in infancy. *J. genet. Psychol.*, 1930, **37**, 485–498.

KASANIN, J. S. *The disturbance of conceptual thinking in schizophrenia*. Berkeley: Univer. California Press, 1944.

KASATKIN, N. I., & LEVIKOVA, A. M. On the development of early conditioned reflexes and differentiation of auditory stimuli in infants. *J. exp. Psychol.*, 1935, **18**, 1–19.

LEUBA, C. Images as conditioned sensations. *J. exp. Psychol.*, 1940, **26**, 345–351.

LEUBA, C. The use of hypnosis for controlling variables in psychological experiments. *J. abnorm. soc. Pschol.*, 1941, **36**, 271–274.

LOOMIS, A. M. A technique of observing the social behavior of nursery school children. *Child develpm. Monogr.*, 1931, No. 5.

MATEER, F. *Child behavior*. Boston: Badger, 1918.

MAYS, L. L. Studies of catatonia: V. Investigation of the perseverational tendency. *Psychiat. Quart.*, 1934, **8**, 728.

MEDNICK, S. A. Distortions in the gradient of stimulus generalization related to cortical

brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51**, 536–542.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316–327.

MEDNICK, S. A., & LEHTINEN, L. F. Stimulus generalization as a function of age in children. *J. exp. Psychol.*, 1957, **53**, 180–183.

MUNROE, RUTH. *Schools of psychoanalytic thought.* New York: Dryden, 1955.

PAINTAL, A. S. A comparison of the GSR in normals and psychotics. *J. exp. Psychol.*, 1951, **41**, 425–428.

PFAFFMAN, C., & SCHLOSBERG, H. The conditioned knee jerk in psychotic and normal individuals. *J. Psychol.*, 1936, **1**, 201–206.

POLLACK, R. H. A genetic study of intuitive word meanings. Unpublished doctoral dissertation, Clark University, 1953.

RAPAPORT D. The conceptual model of psychoanalyses. *J. Pers.*, 1951, **20**, 56–81. (a)

RAPAPORT, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. (b)

RAPAPORT, D., GILL, M., & SHAFER, R. *Diagnostic psychological testing.* Vol. 1. Chicago: Year Book, 1945.

RAZRAN, G. Conditioned responses in children. *Arch. Psychol., NY*, 1933, No. 148.

RAZRAN, G. Conditioned responses: An experimental study and a theoretical analysis. *Arch. Psychol., NY*, 1935, No. 191.

REICHARD, SUZANNE, SCHNEIDER, MARION, & RAPAPORT, D. The development of concept formation in children. *Amer. J. Orthopsychiat.*, 1944, 14, 156–161.

REISS, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, **36**, 143–152.

SCOTT, H. D. Hypnosis and the conditioned reflex. *J. gen. Psychol.*, 1930, **4**, 113–130.

SHIPLEY, W. C. Studies of catatonia: VI. Further investigation of the perseverative tendency. *Psychiat. Quart.*, 1934, **8**, 736–744.

SIEGEL, E. L. Genetic parallels of perceptual structuralization in paranoid schizophrenia: An analysis by means of the Rorschach test. *J. proj. Tech.*, 1953, **17**, 151–161.

SPENCE, K. W., & TAYLOR, J. A. The relation of conditional response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–277.

TAYLOR, J. A., & SPENCE, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–503.

TERMAN, L. M. *The measurement of intelligence.* Boston: Houghton Mifflin, 1916.

WERNER, H. *Comparative psychology of mental development.* New York: Harper, 1940.

WERNER, H. Genetic experimental psychology. In P. L. Harriman (Ed.), *Encyclopedia of psychology.* New York: Philosophical Library, 1946.

WERNER, H., & KAPLAN, B. The developmental approach to cognition: Its relevance to the psychological interpretation of anthropological and ethnolinguistic data. *Amer. Anthropologist*, 1956, **58**, 866–880.

WERNER, H., & KAPLAN, EDITH. Development of word meaning through verbal context: An experimental study. *J. Psychol.*, 1950, **29**, 251–257.

WERNER, H., & KAPLAN, EDITH. *The acquisition of word meanings: A developmental study.* Evanston, Ill.: Child Development, 1952.

WOLFF, W. *The expression of personality.* New York: Harper, 1943.

YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock Deterioration Test. *Psychol. Rev.*, 1941, **48**, 261–267.

ZALUZHNI, A. S. Collective behavior of children at preschool age. *J. soc. Psychol.*, 1930, **1**, 367–378.

(Received November 3, 1960)

Psychological Bulletin
1962, Vol. 59, No. 1, 70–73

THE RELIABILITY OF A RESPONSE MEASURE: DIFFERENTIAL RECOGNITION-THRESHOLD SCORES

DONN BYRNE AND JOAN HOLCOMB
University of Texas *San Francisco State College*

The problem of reliability of measurement is a familiar one in the context of test construction and test evaluation. In other types of investigation, however, responses frequently are measured in a variety of ways by a variety of scoring procedures (often a priori ones) without evident concern about measurement. Although psychometric issues appear to be foreign to experimental methodology, any specified set of stimuli may be conceptualized as a test and the quantification of subjects' responses as test scores. Viewed in this way, such scores should be evaluated according to accepted standards for psychological tests (American Psychological Association, 1954).

An investigation of the validity of a response measure is usually implied in the design of an experiment; reliability, however, is often ignored. Whether results are positive or negative with respect to one's hypotheses, reliability of measurement can assume great importance. Psychologists have the disconcerting tendency to create a new methodology for each experiment. In work on perceptual defense, for example, it is difficult to find any two studies in which the same stimulus is presented in the same way to evoke responses which are quantified in the same manner. It seems reasonable to hypothesize that some of these stimuli presented to subjects in a particular way are going to yield more reliably measured response dimensions than others. With a heterogeneous methodology and unknown reliability coefficients, it should not be surprising to find some degree of inconsistency across experiments not attributable to theoretical weaknesses. Thus, generally positive results mixed with some negative results may reflect differentially reliable "tests." Even the positive results are no guarantee that reliable measures have been employed. McNemar (1960) suggests several factors which act to make published results more likely to involve a false rejection of the null hypothesis than the .05 level of significance would suggest. In addition, it would seem logical to construct reliable measuring techniques as a preliminary step in experimental work rather than as an afterthought.

Differential Recognition Thresholds

As an example of inadequate measurement techniques, some of the "new look" experiments in perception of the past decade will be briefly reviewed. In studies of perceptual defense, differential recognition thresholds for emotionally toned vs. neutral stimulus material have frequently served as the dependent variable and as a measure of individual differences in defensiveness.

All four types of reliability should be considered in utilizing a differential recognition-threshold score. First, if any subjectivity is involved in the scoring process, there should be some determination of the extent to which independent judges are able to arrive at approximately identical scores. Interscorer agreement is a

necessary, but not sufficient, condition for reliability of measurement. Unfortunately, many investigators determine only the reliability of the scoring procedure rather than of the scores themselves. Second, if a series of discrete, presumably homogeneous responses are combined to form a total score, it is important to determine the extent to which this score is internally consistent. Third, if the score is considered to be indicative of an enduring personality characteristic, it is essential to know the extent to which this score is stable over time. Fourth, if a different but theoretically equivalent set of stimuli is employed to elicit responses, the equivalence of the two sets of scores should be determined.

A review of perceptual defense and related studies which have used a differential recognition threshold suggests that a thorough examination of reliability is unusual. As might be anticipated, the reliability coefficient which is most frequently reported is that of interscorer consistency, and the results are generally quite good (Eriksen, 1951a, 1951b; Kogan, 1956; Lazarus, Eriksen, & Fonda, 1951; Stein, 1953). Internal consistency is less frequently investigated, and the reported coefficients range from good (Vanderplas & Blake, 1949) to mediocre (McClelland & Liberman, 1949) to unsatisfactory (Eriksen, 1951a, 1951b). Holtzman and Bitterman (1956) reported that perceptual thresholds for taboo and neutral words were unreliable measures; therefore, they were eliminated from a factor analytic study. An investigation of the stability of differential recognition thresholds over time was not reported in any of the studies reviewed. Stein's (1953) data indicate that equivalent forms of the stimulus which he used yielded very similar results. The majority of the studies using a differential threshold score as a variable report no reliability information (Beier & Cowen, 1953; Carpenter, Wiener, & Carpenter, 1956; Chodorkoff, 1954; Cowen & Obrist, 1958; Greenbaum, 1956; Kissen, Gottesfeld, & Dicks, 1957; Kurland, 1954; Postman & Brown, 1952; Smith, 1954; Spence, 1957; Wiener, 1955; Zuckerman, 1955).

An Unreliable Score

The senior author planned to use the differential recognition threshold for hostile vs. neutral words presented tachistoscopically as a criterion measure for a new test designed to measure repressing and sensitizing defenses. It should be confessed that the reliability investigation was undertaken only when certain difficulties were encountered.

Twenty pairs of hostile and neutral words were each matched for length, initial letter, and frequency of occurrence in one million words according to the Thorndike-Lorge (1944) word count. Hostile words were defined as those representing behavior involving the derogation, injury, or destruction of either animate or inanimate objects. Neutral words were defined as those which were not emotionally toned. A word was assigned to either category on the basis of the unanimous agreement of three independent judges. The 40 words were placed on slides and arranged in random order.

The slides were used with a Keystone Overhead Projector equipped with a Flashometer. Following a demonstration with a neutral practice word, each stimulus word was presented at 1/100, 1/75, 1/50, 1/37.5, 1/25, 1/10, and 1 second. After each trial, the subject responded by writing down his best guess as to the word presented. Subjects were seen in small groups.

The threshold for each word consisted of the first trial on which that word was correctly recognized. Scores ranged from 1 (correct recognition at the 1/100 presentation) through 8 (failure to recognize the word on any trial). A subject's mean threshold on the 20 neutral words minus his mean threshold on the 20 hostile words yielded a defense score. Presumably, a positive score would indicate a sensitizing reaction and a negative score a repressing reaction.

Disappointing results in cross-validating the test that was being developed led to a belated investigation of the reliability of the criterion. Differential thresholds were obtained for almost 600 subjects, men and women enrolled in general education courses at San Francisco State College. From this total, a sample of 50 was drawn. Because some subjectivity enters into the determination of the trial on which correct recognition first occurs, the authors scored these protocols independently. The defense scores had respectable interscorer consistency as shown by a correlation of .91.

The second type of reliability considered was internal consistency. When differential thresholds are obtained, it is assumed that there is some response homogeneity with respect to stimulus content. In this study, responses to 20 of the projected words should have been determined in part by their common reference to hostility, and these responses should to some extent differ from those evoked by the 20 nonhostile words. Therefore, split-half reliability was determined by dividing the hostile words into odd and even groups and computing the differential threshold scores for these two groups compared to their matching neutral words. The coefficient of internal consistency was .00. It was not deemed essential to apply the Spearman-Brown correction formula.

Thus, independent judges agreed about the nature of the stimulus material and about the scoring of the subjects' responses. Nevertheless, the resulting scores were unreliable. In view of this finding and the considerations discussed earlier, it is suggested that, whenever possible, any study should include a report of the reliability of its response measures.

REFERENCES

AMERICAN PSYCHOLOGICAL ASSOCIATION, Committee on Psychological Tests. *Technical recommendations for psychological tests and diagnostic techniques.* Washington, D. C.: APA, 1954.

BEIER, E. G., & COWEN, E. L. A further investigation of the influences of "threat-expectancy" on perception. *J. Pers.*, 1953, **22**, 254–257.

CARPENTER, B., WIENER, M., & CARPENTER, JANETH T. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 380–383.

CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49**, 508–512.

COWEN, E. L., & OBRIST, P. A. Perceptual reactivity to threat and neutral words under varying experimental conditions. *J. abnorm. soc. Psychol.*, 1958, **56**, 305–310.

ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564. (a)

ERIKSEN, C. W. Some implications for TAT interpretation arising from need and perception experiments. *J. Pers.*, 1951, **19**, 282–288. (b)

GREENBAUM, M. Manifest anxiety and tachistoscopic recognition of facial photographs. *Percept. mot. Skills*, 1956, **6**, 245–248.

HOLTZMAN, W. H., & BITTERMAN, M. E. A factorial study of adjustment to stress. *J. abnorm. soc. Psychol.*, 1956, **52**, 179–185.

KISSEN, B., GOTTESFELD, H., & DICKS, R. Inhibition and tachistoscopic thresholds for sexually charged words. *J. Psychol.*, 1957, **43**, 333–339.

KOGAN, N. Authoritarianism and repres-

sion. *J. abnorm. soc. Psychol.*, 1956, **53**, 34–47.

KURLAND, S. H. The lack of generality in defense mechanisms as indicated in auditory perception. *J. abnorm. soc. Psychol.*, 1954, **49**, 173–177.

LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.

MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition for need-related words. *J. Pers.*, 1949, **18**, 236–251.

MCNEMAR, Q. At random: Sense and nonsense. *Amer. Psychologist*, 1960, **15**, 295–300.

POSTMAN, L., & BROWN, D. R. The perceptual consequences of success and failure. *J. abnorm. soc. Psychol.*, 1952, **47**, 213–221.

SMITH, J. G. Influence of failure, expressed hostility, and stimulus characteristics on verbal learning and recognition. *J. Pers.*, 1954, **22**, 475–493.

SPENCE, D. P. A new look at vigilance and defense. *J. abnorm. soc. Psychol.*, 1957, **54**, 103–108.

STEIN, K. B. Perceptual defense and perceptual sensitization under neutral and involved conditions. *J. Pers.*, 1953, **21**, 467–493.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1944.

VANDERPLAS, J. M., & BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252–256.

WIENER, M. Word frequency or motivation in perceptual defense. *J. abnorm. soc. Psychol.*, 1955, **51**, 214–218.

ZUCKERMAN, M. The effect of frustration on the perception of neutral and aggressive words. *J. Pers.*, 1955, **23**, 407–422.

(Received November 4, 1960)

Psychological Bulletin
1962, Vol. 59, No. 1, 74–76

COMMENTS ON "THE PARAMORPHIC REPRESENTATION OF CLINICAL JUDGMENT"[1]

JOE H. WARD, JR.
Personnel Laboratory, Wright Air Development Division

The purpose of this discussion is to comment on possible misunderstandings that may arise from the discussion of relative weights presented in Hoffman's (1960) paper, "The Paramorphic Representation of Clinical Judgment." First, the present author certainly agrees that regression techniques can be quite useful in the study and analysis of judgments. Regression analysis can certainly play an important role in the study of the homogeneity of judgment policies among individuals and in the analysis of the extent to which variables contribute to judgment.

The relative weights presented on page 120 in Hoffman's article may lead to a certain amount of misunderstanding about the "independent contribution" of a variable in the judgment process. Before discussing this point it is necessary to establish what is meant by the term "independent contribution" of a variable.

Consider a set of variables, X_1 $X_2, \cdots X_k$ which all have mean values equal to 0. The independent contribution to prediction of Y of a single predictor, say X_1, refers to the amount of predictive efficiency that the residual vector E in the vector X_1 can make when predicting the criterion Y, where the residual E refers to the error remaining when X_1 is predicted from a least squares combination of $X_2, X_3, \cdots X_k$.

That is, if:

$$X_1 = w_2X_2 + w_3X_3 + \cdots + w_kX_k + E$$

where values for w_i $(i = 2, \cdots k)$ are "least squares" coefficients; and if:

$$Y = b_1E + G$$

where b_1 is determined by least squares and G is the error remaining when Y is predicted from E, then the idea of independent contribution refers to the extent to which the residual E can account for the criterion Y. Frequently the term independent contribution refers to the proportion of the total variance of Y that is accounted for by the residual in X_1.

A regression coefficient reflects the value of the independent contribution *only* when the regression coefficient equals zero—and then, of course, there is no independent contribution. When a particular regression coefficient is different from zero, very little can be said about the independent contribution that the particular variable associated with the coefficient makes toward prediction of the independent variable.

The concept of relative weight might lead to some confusion about the independent contribution of the corresponding predictor. Not only does it seem difficult to attach meaning to positive nonzero relative weights but it seems particularly difficult to interpret negative relative weights.

Consider first the specific example that Hoffman presents on page 122, in which it is assumed that $r_{01} = .400$, $r_{02} = .000$, $r_{12} = .707$. The solution of the matrix is indicated to yield $\beta_{01} = .800$, $\beta_{02} = -.566$, $R_{0.12} = .566$, and squaring the value of $R_{0.12}$ we

[1] The research reported in this paper was sponsored by Personnel Laboratory, Wright Air Development Division, under Research and Development Project 7719, Task 17112.

obtain $R^2_{0.12}=.32$. Since the square of the correlation of a predictor with the judgment criterion indicates the proportion of variance accounted for when no other variables are considered, we can observe that the proportion of variance accounted for by Predictor Number 1 is .16. It is quite apparent that the second predictor, when predicting alone, accounts for no variance in the criterion. However, let us see what its independent contribution is. The $R^2_{0.12}$ of the least-squares combination is .32; therefore, even though the second predictor accounts for no variance when predicting alone, its independent contribution is equal to 16% (.32−.16) of the total criterion variance. From this it becomes apparent that, whereas the first predictor when used alone can account for only 16% of the total criterion variance, the use of the second predictor provides an additional 16% of the variance. Furthermore, we can see that the independent contribution of Predictor Number 1, is 32% (.32−.0) of the total criterion variance.

Several more examples are presented in Table 1. The columns of Table 1 are defined as:

r_{01} = correlation between Predictor 1 and the criterion

r_{02} = correlation between Predictor 2 and the criterion

r_{12} = correlation between Predictor 1 and Predictor 2

β_{01} = standardized regression coefficient for Predictor 1

β_{02} = standardized regression coefficient for Predictor 2

$R^2_{0.12}$ = squared multiple correlation resulting from prediction by Variables 1 and 2

$R^2_{01.2}=R^2_{0.12}-r^2_{02}$ = the independent contribution of Predictor 1

$R^2_{02.1}=R^2_{0.12}-r^2_{01}$ = the independent contribution of Predictor 2

w_{01} and w_{02} = relative weight for predictors (see Hoffman, 1960, p. 122)

Table 1 reveals some of the difficulty of using the idea of relative weight in the interpretation of contributions of individual variables. It can be observed, for example, that the relative weights in several different problems can be identical while the independent contributions can be quite different. Evidently, the concept of relative weight will not pro-

TABLE 1

EXAMPLES OF SOLUTIONS TO PREDICTOR REGRESSION PROBLEMS

Example No.	r_{01}	r_{02}	r_{12}	β_{01}	β_{02}	$R^2_{0.12}$	Independent contributions		Relative weights	
							$R^2_{01.2}$	$R^2_{02.1}$	w_{01}	w_{02}
1	.400	.000	.707	.800	−.566	.320	.320	.160	1.000	.000
2	.400	.000	.100	.404	−.004	.162	.162	.002	1.000	.000
3	.800	.000	.100	.808	−.081	.646	.646	.006	1.000	.000
4	.800	.000	.600	1.250	−.750	1.000	1.000	.360	1.000	.000
5	.700	.000	.700	1.373	−.961	.961	.961	.471	1.000	.000
6	.200	.700	.800	−1.000	1.500	.850	.360	.810	−.235	1.235
7	.100	.800	.500	−.400	1.000	.760	.120	.750	−.053	1.053
8	.998	.000	.070	1.002	−.702	1.000	1.000	.002	1.000	.000
9	.707	.000	.707	1.414	−1.000	1.000	1.000	.500	1.000	.000
10	.070	.000	.998	14.29	−14.27	1.000	1.000	.995	1.000	.000

vide any information about the independent contribution of a predictor. In general, the relative weight will not reveal anything about the independent contributions of the predictors except in the special case of orthogonality among the predictors.

REFERENCE

HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, **57**, 116–131.

(Received December 26, 1960)

Psychological Bulletin
1962, Vol. 59, No. 1, 77–80

ASSESSMENT OF THE INDEPENDENT CONTRIBUTIONS OF PREDICTORS[1]

PAUL J. HOFFMAN
Oregon Research Institute, Eugene

The term "independent," like so many others in the jargon, is cursed by having acquired a number of meanings. Perhaps its usage could be restricted to imply only experimental independence, and the term "orthogonal" to convey the notion of statistical independence. Then both Ward (1962) and I could be wrong together. But as things stand, I believe that the differences between us are mostly semantic, and therefore minor.

Ward's "independent contribution" indicates the proportion of variance in the criterion attributed to the residual in a predictor, X_1, after variance common to X_1 and other predictors is removed. Such coefficients are heavily dependent upon the interrelations among the variables included for analysis. Precisely, the independent contribution of X_1 will necessarily be reduced by another predictor, X_2, and by an amount equal to the variance common to X_1, X_2, and the criterion, X_0.

Let $r^2_{0(1.)}$ be the independent contribution of X_1 *before* the inclusion of X_2 and $r^2_{0(1.2)}$ be its reduced independent contribution. Then if we express the amount of this reduction by θ and the percent reduction by $\hat{\theta}=\theta/r^2_{0(1.)}$, it can be shown that:

$$\theta = r^2_{01} + r^2_{02} - R^2_{0(1,2)}$$

$$\hat{\theta} = \frac{r^2_{01} + r^2_{02} - R^2_{0(1,2)}}{r^2_{01}}$$

It follows that although X_1 be a very satisfactory test of high validity, it will yield no independent contribution if all of its valid variance may be predicted from a linear combination of a set of other predictors. This of course holds even if the set of other predictors contains no variable which is individually as predictive as X_1. Despite this limitation, coefficients of independent contribution are quite useful, especially in empirical prediction studies.

The problem of assessing the relative contribution of a variable, i.e., the relative importance of that variable as compared with others included with it in the same set, is different from the problem of prediction. It is different because the primary concern is that of determining some mathematical representation of relative importance. This may be achieved through some kind of partitioning of the criterion variance.

The variance of predicted scores may be partitioned in many ways, but few are psychologically meaningful. Apportioning it among beta coefficients or squared beta coefficients is not meaningful, since not all of the predictable variance is accounted for. Thus, in the two-predictor case, using McNemar's (1955) notation, the variance of predicted scores $\sigma^2_{Z'_1}$ may be expressed as follows:

$$\sigma^2_{Z'_1} = \beta^2_2 + \beta^2_3 + 2\beta_2\beta_3 r_{23}$$

When $r_{23} \neq 0$, the squared betas cannot account for the predictable criterion variance exclusively in

[1] This investigation was supported by a United States Public Health Service Research Grant (M2097-C1) from the National Institute of Mental Health.

terms of the independent contributions of the predictors for the simple reason that there exists a *joint* contribution as well. This joint contribution may be more than modest, particularly where the number of predictors is large and their intercorrelations at least moderate. Beta coefficients are therefore inadequate simply because no linear combination of beta coefficients or of their squares exists which will unambiguously account for the predictable variance of judgments. The same may be said of Ward's independent contribution coefficients. The only exception is the special case in which the predictors are orthogonal. On the other hand, relative weight: defined as:

$$w_{oi} = \frac{\beta_{oi} r_{oi}}{R^2_{0.12\cdots k}} \qquad [1]$$

provides a means of portraying the relative contributions of each of the predictors such that a simple sum of them accounts entirely and unambiguously for the predictable variance. For a contrary point of view on this problem the reader may wish to refer to Ezekiel (1930).

The concept of "relative weight" was developed to provide a means by which the cognitive processes of clinicians (and, for that matter, anyone making judgments or decisions) might be described. It should be noted in this connection that the problem of describing the judgment process differs from the problem of prediction in another way. Judgment studies of the type described in my previous *Psychological Bulletin* article (Hoffman, 1960) deal with a system of variables which is finite or "closed." Since in the experimental arrangement by which the judgments are obtained only known quantitative information is available to the judge, the criterion variance (variance of judgments) must be completely accounted for by two factors: one of these involves some combination of the predictor information, not necessarily linear, perhaps quite complex; the other factor is chance. This being the case, it is meaningful to speak of the possibility of measuring the degree to which the criterion variance may be accounted for by the relation of one variable to the others available to the judge, i.e., within the system but completely independent of external variables which might be thrown into the regression analysis at will.

A fair test of a coefficient which presumably reflects the relative contribution of a predictor in the judgment process is one which compares the value of the obtained coefficient when the predictor is a member of the set available to the judge with its value when the predictor is absent but yet included in the multiple regression analysis. A predictor available to the judge and "used" by him should be capable of being described by a coefficient which has at least a moderate value. When this predictor is experimentally absent from the judgment situation, the value of the coefficient should drop to a chance level. A poor (in this sense) type of coefficient would be one which is affected little by this kind of manipulation.

The coefficient which Ward refers to as independent contribution will not ordinarily satisfy this test. The introduction of an external predictor into such a closed system reduces the independent contributions of the internal predictors since variance which is common to an original predictor, an external predictor, and

the criterion is subtracted. Not everything is bad, however. The independent contribution of such an external predictor may be expected to be no greater than chance, and this will assuredly be reflected in the near-zero values of the coefficient. But the values of the coefficients for the original predictors will nevertheless be determined by the interrelationships among the variables included in the analysis. The coefficient of independent contribution is therefore unsuitable for this purpose. Beta coefficients are likewise affected by such manipulations although not to such a great extent. On the other hand, relative weights pass this test with flying colors, as will be shown in a forthcoming article.

Another point raised by Ward has to do with the relationship between relative weights and coefficients of independent contribution. Although the concepts are different, I do not completely agree that no relationship exists between relative weights and independent contributions. It may be shown, for example, that in the case of two predictors, multiplying the ratio of the two independent contributions by the ratio of the corresponding squared validity coefficients yields the ratio of the squared relative weights. That is:

$$\frac{w^2_{01}}{w^2_{02}} = \frac{r^2_{0(1.2)}}{r^2_{0(2.1)}} \cdot \frac{r^2_{01}}{r^2_{02}} \qquad [2]$$

It is of interest also to note that the ratio of the independent contributions is equal to the ratio of the squared beta coefficients. Multiplying Equation 2 by $\beta^2_{01}/\beta^2_{02}$ we obtain:

$$\frac{w^2_{0}\beta^2_{01}}{w^2_{02}\beta^2_{02}} = \frac{r^2_{0(1.2)}}{r^2_{0(2.1)}} \cdot \frac{r^2_{01}\beta^2_{01}}{r^2_{02}\beta^2_{02}} \qquad [3]$$

and since, from Equation 1:

$$r^2_{oi}\beta^2_{oi} = w^2_{oi} \cdot R^2_{0(123\cdots k)}$$

it follows that:

$$\frac{r^2_{0(1.2)}}{r^2_{0(2.1)}} = \frac{\beta^2_{01}}{\beta^2_{02}} \qquad [4]$$

In contrast, relative weights may be said to represent a product of two proportions: the first is the proportion that the independent contribution bears to the residual of the predictor when the joint effects of the other predictors have been removed. This term is identically equal to the squared beta coefficient. The second term is the proportion of total predictable variance in the criterion which is common to the predictor in question.

To conclude, my usage of the phrase independent contribution is intended to connote that the variance of predicted scores may be successfully partitioned into a simple sum of ingredients, each referring to a specific predictor and each being independent of any joint effect or interaction. What Ward means by the term independent contribution is more ordinarily known as the "part correlation" or, actually, the square of the part correlation (cf. DuBois, 1957). Using the term in Ward's sense, I thoroughly agree with him that the concept of relative weight provides little information about the independent contribution of a predictor. It was not intended to provide such information. It is also correct that the concept of independent contribution provides little information about the contribution of a predictor relative to the contributions of other predictors in a given set, nor is it intended as a method for assessing this aspect of the judgment process.

REFERENCES

DuBois, P. H. *Multivariate correlational analysis.* New York: Harper, 1957.

Ezekiel, M. *Methods of correlation analysis.* New York: Wiley, 1930.

Hoffman, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, **57**, 116–131.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1955.

Ward, J. H., Jr. Comments on "The paramorphic representation of clinical judgment." *Psychol. Bull.*, 1962, **59**, 74–76.

(Received January 21, 1961)

ERRATA

In the article by P. L. Broadhurst and J. L. Jinks in the September 1961 issue, the second equation at the top of the second column on page 338 should read:

$$[h] = \overline{F}_1 - 4\overline{F}_2 - \tfrac{1}{2}\overline{P}_1 - \tfrac{1}{2}\overline{P}_2 + 2\overline{B}_1 + 2\overline{B}_2$$

The first equation in the second column on page 353 should read:

$$[h] - [i] = \overline{F}_1 - \tfrac{1}{2}(\overline{P}_1 + \overline{P}_2)$$

In the article by Charles S. Morrill in the September 1961 issue, the references by R. E. Silverman were published by the United States Naval Training Device Center. The reference by N. A. Crowder, 1959b, was published in 1955.

In the article by Mark R. Rosenzweig in the September 1961 issue, the quotation from Ades and Brookhart on page 384 should read as follows: "that the inferior colliculus with its strong commissural connections and connections to efferent [not *a*fferent] mechanisms may be the principal device responsible for localization" (1950, p. 203).